nature climate change

Brief Communication

https://doi.org/10.1038/s41558-024-01943-1

Offshoring emissions through used vehicle exports

Received: 19 July 2023

Accepted: 26 January 2024

Published online: 20 February 2024

Check for updates

Saul Justin Newman[1,2] ✉, Kayla Schulte[1,3], Micol Matilde Morellini[1,4,5], Charles Rahal[1,4] & Douglas R. Leasure[1,4]

Policies to reduce transport emissions often overlook the international flow of used vehicles. We quantify the rate at which used vehicles generated CO_2 and pollution for all used vehicles exported from Great Britain—a globally leading used vehicle exporter—across 2005–2021. Destined for low–middle-income countries, exported vehicles fail roadworthiness standards and, even under extremely optimistic 'functioning-as-new' assumptions, generate at least 13–53% more emissions than scrapped or on-road vehicles.

Transport is the largest emitting sector of greenhouse gases, accounting for a quarter to a third of all emissions in developed countries[1,2] with serious consequences for both climate and health[3–5]. Air pollutants such as nitrogen oxides (NO_x), which are effectively reduced when standards are enforced, cause millions of deaths each year[4,5]. These impacts fall unequally on lower–middle-income countries (LMICs)[6], which suffer more overall and per capita pulmonary deaths from air pollution[4,7] and stand to suffer the greatest impacts from climate change[8].

The source of vehicles in LMICs is dominated by unregulated trade[1–3,9,10]. As of 2020, 100 countries receiving used vehicles had no vehicle emissions standards[9] and only 11 had 'very good'[9] emissions regulations. However, the United States, European Union, Japan and United Kingdom collectively supply 90% of used vehicles exported to (non-EU) LMICs[9]. The potential for rapid regulation is therefore incumbent on just four jurisdictions, all of which already maintain high vehicle emissions standards.

Using comprehensive government databases[11], we quantify per kilometre rates at which vehicles generate carbon and pollution for every vehicle ($N = 6{,}921{,}292$) legally exported from Great Britain between January 2005 and December 2021. We compare these vehicle emissions to every private vehicle driven in Great Britain during the same period and those that would have been driven if they had not been scrapped.

These data reveal substantially higher rates of CO_2 and pollution generation in exported vehicles, even under optimistic 'functioning-as-new' emissions intensity estimates that assume no vehicle modifications or vehicle degradation with age (Fig. 1 and Extended Data Fig. 1). Exported cars generate at least 23 g (13%) more CO_2 per kilometre than cars scrapped in the same period (Fig. 1a and Supplementary Code) and at least 29 g (17%) more CO_2 per kilometre than the contemporary on-road used car fleet (Fig. 1a; mean 197.0, 174.4 and 168.6 g km^{-1} CO_2 for exported, scrapped and on-road fleets, respectively; interquartile ranges (IQRs) 170.1–225.3, 147.5–189.1 and 134.2–188.9 g km^{-1}).

Emissions figures were even more striking for other pollutants. Exported cars emit similar amounts of hydrocarbon particulates (Fig. 1b) but 48 mg km^{-1} more NO_x (53% higher; Fig. 1c) than scrapped cars. Likewise, observed engine capacities were larger (Fig. 1d) and fuel efficiency at least 9% worse, by 3.3 miles per gallon (MPG, mean 38.5, 41.8 and 44.4 mpg for exported, scrapped and on-road used fleets respectively; IQR 33–45 mpg exported, 37–47 mpg scrapped and 37–49 mpg on-road fleets).

A substantial fraction (42%) of exported diesel vehicles were predicted to fail the current EURO-4 emissions standards[12] that form the legal roadworthy minimum for all vehicles registered after 2000. A surprising 83% were predicted to fail the EURO-6 diesel[12] CO_2 emissions standards, and 98% failed the EURO-6 carbon monoxide and NO_x standards. These differences are not the result of overdispersion, where a few high-emitting exports[12] dragged up the average; similar or even larger gaps in pollution rates were observed for the median pollution rates of exported, scrapped and contemporary on-road used vehicles (Fig. 1).

Daily-resolution data for six million exported cars reveals that the gap between exported, on-road and scrapped fleets is consistent over time (Fig. 2a), apart from a narrowing and then rapid expansion of this gap over 2020–2021 alongside distortions of trade patterns, used car prices and vehicle testing regulations during the COVID-19 pandemic (Fig. 2b). That is, Great Britain persistently scraps lower-emissions vehicles while exporting higher-emission vehicles (Fig. 2b). Whereas geographical disparities due to uneven

[1]Leverhulme Centre for Demographic Science, Nuffield Department of Population Health, University of Oxford, Oxford, UK. [2]University College, University of Oxford, Oxford, UK. [3]MRC Centre for Environment and Health, Imperial College London, London, UK. [4]Nuffield College, University of Oxford, Oxford, UK. [5]Department of Sociology, University of Oxford, Oxford, UK. ✉e-mail: saul.j.newman@gmail.com

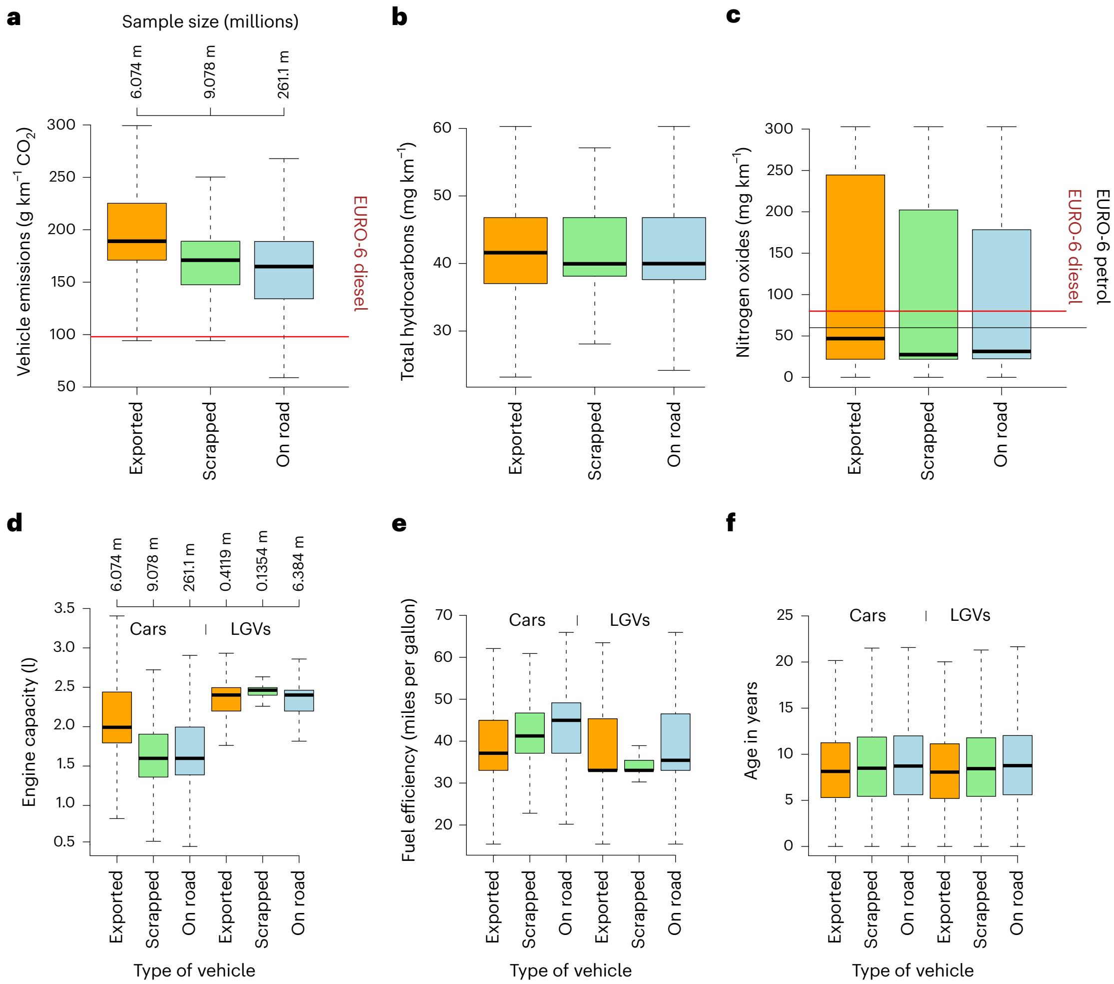


Fig. 1 | Pollution generated per kilometre by used vehicle fleets. **a**–**c**, Exported cars (orange) generate at least 13–24% more CO_2 per kilometre (**a**), similar average but (through overdispersion) higher median fine particulate matter pollution (**b**) and at least 53% more nitrogen oxides per kilometre than scrapped (green) or on-road (blue) vehicle fleets (**c**). **d**–**f**, Both exported car and light goods vehicle (LGV) fleets also have larger observed engine capacities (**d**) and worse fuel efficiency (**e**) despite being similar ages to on-road fleets and much younger than contemporary scrapped vehicles (**f**). Boxes show IQRs, whiskers are 1.5× IQR, sample size (millions) shown on upper *x* axes, summary data supplied in a Figshare repository[25].

concentrations of upmarket export vehicles or differing 'on-road' usage were anticipated, this export gap was remarkably uniform. Almost every British postcode region (95%), representing a full cross section of society, export higher-polluting vehicles than those they drive or scrap (Fig. 2c).

Exported vehicles will probably generate more pollution per kilometre independent of their destinations and patterns of use for simple physical reasons: compared to their more-efficient scrapped alternatives, export vehicles have larger observed engine capacities (Fig. 1d) and lower operating efficiencies (Fig. 1e), despite a younger average age (Fig. 1f). These fixed factors also mean that degradation rates are probably rank conserved, especially for CO_2, and high-polluting vehicles will remain the most polluting as vehicle fleets age.

Air quality outcomes are more nuanced than individual emissions categories, including how vehicles are driven, road conditions, engine age, climate, payload and maintenance schedules[13,14]. Vehicles also generate pollution, such as ozone or non-tailpipe emissions, for which testing data were not available. Observational data are urgently needed to fill this gap.

Adding to the challenges of measuring emissions, emissions testing data have long been manipulated—for example, during 'Dieselgate', where nine major manufacturers used 'defeat devices' to alter performance and intentionally deceive environmental agencies and regulators. In the Dieselgate aftermath, vehicle manufacturers are, incredibly, allowed to legally manipulate vehicles during new car emissions testing[14,15] by, for example, removing wing mirrors and seats, taping up high-drag surfaces or hard-baking and over-inflating tyres. Manufacturers are also now allowed to 'adjust' emissions estimates[15] by 4.5% and programme vehicles to turn off emissions-reduction devices[16] when the weather becomes 'too hot' or 'cold'[16]. Manufacturers define 'hot' and 'cold'. For example, Renault told the French government their emissions control devices should shut off above 35 °C and below 17 °C (Paris is colder 83% of the time[15]) to 'protect the engine'[15], at the cost of protecting the climate and human health. As a result, real-world emissions increasingly overshoot (currently by ~50%) emissions measured during testing[14,17,18]. Some 13% of diesel cars in the European Union now emit NO_x at over ten times the legal standard[18], outnumbering the 10% that actually meet those standards.

The lack of emissions standards in most destination countries also results in the routine stripping of emissions-reduction devices for resale[19] or melting down before export[9,19]. One study tested 160 vehicles destined for Africa from the European Union[19]. Of the vehicles that could start, 85–93% failed to meet the (roadworthy minimum) EURO-4 emissions standards[20], 20% of petrol (gasoline) vehicles did not comply with any emissions standard at all[19] and 10% had their catalytic converters cut out[19], increasing NO_x and carbon monoxide pollution tenfold.

 https://doi.org/10.1038/s41558-024-01943-1

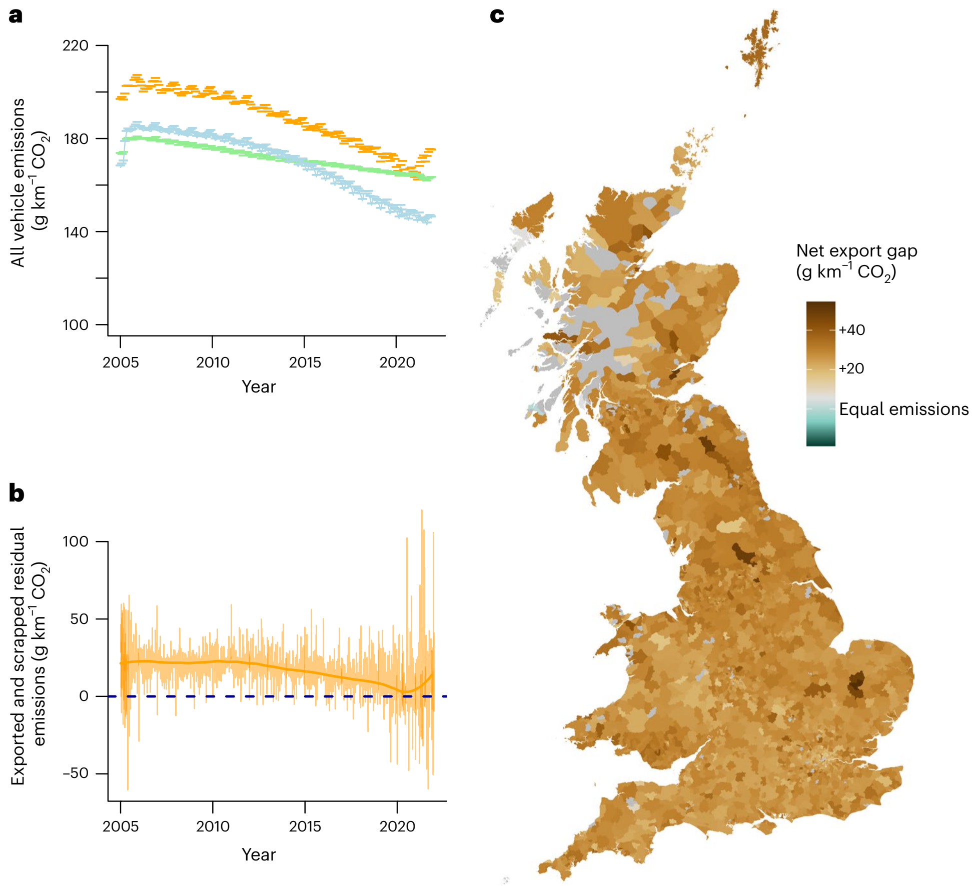


Fig. 2 | Export of consistently higher-emitting vehicles from Great Britain. **a**, Exported vehicles (orange) are consistently more polluting than contemporary scrapped (green) and used on-road (blue) vehicles. **b**, Cleaner choices and improving standards have lowered exported emissions after considerable delay but the gap between exported and scrapped fleets persists and is growing post-pandemic. **c**, Scrapped vehicles are cleaner than exports across 5,874 (95%) of our 6,145 observed postcode regions (2005–2021 inclusive), covering every community in Britain. Grey regions have insufficient data or no inspection sites. Orange line in **b** shows locally weighted smoothed spline; note non-zero *y* axis in **a** to emphasize variation.

Used vehicles were exported at an average 8.5 years of age (IQR 5.0–10.6 years), and both emissions[13] and fuel efficiency[14] degrade with age. Our models, therefore, probably underestimate vehicle pollution rates substantially by relying on new car testing data that do not account for increasing emissions generation from vehicle ageing or modifications. The non-stationarity and complexity of emissions degradation curves[13,14] and vehicle modifications mean that direct measurements—such as those increasingly captured by annual vehicle emissions tests required in the United Kingdom—are needed to improve estimates of exported emissions. Actual emissions are probably far higher[13,14,18], perhaps 150% higher for CO_2 under ideal 'European-style' driving conditions[18], but enormous and unnecessary gaps in our knowledge remain.

As with carbon leakage from heavy industry and manufacturing[6], rich countries appear to be offshoring the cost of replacing high-polluting vehicles. There are, however, some positive trends. Whereas many improved emissions are artefacts of manipulated testing[16,17,21], better fuel efficiency and air quality standards in the United States, United Kingdom, European Union and Japan are slowly reducing the estimated pollution of exported vehicles over time (Fig. 2a,b for Great Britain). These four jurisdictions are the collective source of over 95% of light used vehicle exports worldwide and despite creating 40% of global transport emissions[22], implement world-leading vehicle emissions standards inside their own borders.

Imparting the same standards on exported vehicles, preventing the removal of emissions-reduction devices and redirecting clean vehicles from the scrapyard to the export fleet would all positively impact global emissions. Export licences can be indexed to increase duties on dirty vehicles or subsidize clean vehicle exports. Export countries have very few major vehicle ports[1] and thousands of mechanics qualified to evaluate legal roadworthy standards. Tasking mechanics to randomly spot check vehicles in port and issue penalties when vehicles fail emissions tests would be an extremely low-cost intervention to stem the dirty used vehicle trade.

Such measures would not necessitate increased vehicle prices, which can reduce access to the economic benefits of vehicle ownership. Supply shocks can be mitigated or avoided by using policy and incentives and by redirecting clean vehicles from scrapyards to export. Cleaner vehicles also have smaller and more fuel-efficient engines on average and lower ongoing costs[23] over the life of the vehicle, reducing net economic burdens.

Potential short-term price increases imposed on individuals are also offset by long-term reduction in the societal and economic costs from pollution[2,4,5] and climate change[1,3,8]. Most LMICs are placing this consideration above others, with widespread moves to ban the import of dirty used vehicles, regardless of price shocks[9,24]. These policies reflect a growing desire for clean air over cheap cars. However, such moves are struggling for traction due to a lack of policing and resources and the unstemmed flow of unregulated imports[24].

Developed economies can aid these goals and reduce the damage from vehicular emissions by raising export standards to match their own internal legal minimum standards[12,20]. Such low-cost interventions are an immense opportunity for rich high-emitting countries to reduce global emissions and cut pollution in the developing world. To instead overlook this problem and allow the continued flow of high-emissions vehicles would be a devastating missed opportunity and an ethical failure.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41558-024-01943-1.

References

1. *CO_2 Emissions from Fuel Combustion* OECD iLibrary 514 (International Energy Agency, 2019); https://doi.org/10.1787/2a701673-en
2. Wallington, T. J. & Wiesen, P. N_2O emissions from global transportation. *Atmos. Environ.* **94**, 258–263 (2014).
3. Anenberg, S. C. et al. Impacts and mitigation of excess diesel-related NO_x emissions in 11 major vehicle markets. *Nature* **545**, 467–471 (2017).
4. Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D. & Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **525**, 367–371 (2015).
5. Cohen, A. J. et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet* **389**, 1907–1918 (2017).
6. Babiker, M. H. Climate change policy, market structure, and carbon leakage. *J. Int. Econ.* **65**, 421–445 (2005).
7. Fuller, R. et al. Pollution and health: a progress update. *Lancet Planet. Health* **6**, e535–e547 (2022).
8. Mertz, O., Halsnæs, K., Olesen, J. E. & Rasmussen, K. Adaptation to climate change in developing countries. *Environ. Manage.* **43**, 743–752 (2009).
9. United Nations Environment Programme *Used Vehicles and the Environment—A Global Overview of Used Light Duty Vehicles: Flow, Scale, and Regulation* (United Nations, 2020).
10. Davis, L. W. et al. International trade in used vehicles: the environmental consequences of NAFTA. *Am. Econ. J. Econ. Policy* **2**, 58–82 (2010).
11. *Car and Fuel Emissions Information* (Vehicle Certification Agency, 2022); https://carfueldata.vehicle-certification-agency.gov.uk/downloads/default.aspx
12. European Commission. Commission regulation (EU) number 459/2012. *Off. J. Eur. Union* **L**, 142/16 (2012).
13. Bernard, Y., Dallmann, T., Tietge, U., Badshah, H. & German, J. *Development and Application of a US Real-World Vehicle Emissions Database* (The Real Urban Emissions Initiative, 2020); https://www.trueinitiative.org/data/publications/development-and-application-of-a-us-real-world-vehicle-emissions-database
14. Greene, D. L. et al. How does on-road fuel economy vary with vehicle cumulative mileage and daily use? *Transp. Res. Part D Transp. Environ.* **55**, 142–161 (2017).
15. Muncrief, R., German, J. & Schultz, J. *Defeat Devices Under the U.S. and EU Passenger Vehicle Emissions Testing Regulations* (ICCT, 2016); https://policycommons.net/artifacts/10663812/defeat-devices-under-the-us/11571214/
16. European Commission.Guidance on the identification of the presence of defeat devices with regards to emissions of light-duty vehicles approved with real driving emissions (RDE), heavy-duty vehicles and on anti-tampering protection (2023/C 68/01). *Off. J. Eur. Union* **68**, 1–20 (2023).
17. Tietge, U. et al. *A 2015 Update of Official and 'Real-World' Fuel Consumption and CO_2 Values for Passenger Cars in Europe* (The International Council for Clean Transport, 2015).
18. Baldino, C. *Road Tested: Comparative Overview of Real-World Versus Type-Approval NO_x and CO_2 Emissions from Diesel Cars in Europe* (The International Council for Clean Transport, 2017); https://theicct.org/wp-content/uploads/2021/06/ICCT_RoadTested_201709.pdf
19. *Used Vehicles Exported to Africa: A Study on the Quality of Used Export Vehicles* (Netherlands Human Environment and Transport Inspectorate., 2020); https://www.ilent.nl/binaries/ilt/documenten/rapporten/2020/10/26/rapport–used-vehicles-exported-to-africa/RAPPORT-+Used+vehicles+exported+to+Africa.pdf
20. European Commission. Directive 98/69/EC of the European Parliament and of the council.*Off. J. Eur. Commun.* **L**, 350/1 (1998).
21. Reynaert, M. *Self Regulation, Corrective Policy and Goodhart's Law: The Case of Carbon Emissions from Automobiles* (NBER, 2016); http://www.nber.org/papers/w22911
22. Crippa, M. et al. *CO_2 Emissions of All World Countries—JRC/IEA/PBL 2022 Report* EUR 31182 EN (Publications Office of the European Union, 2022); https://doi.org/10.2760/730164
23. Miotti, M. et al. Personal vehicles evaluated against climate change mitigation targets. *Environ. Sci. Technol.* **50**, 10795–10804 (2016).
24. Ayetor, G. K. et al. Vehicle regulations in Africa: impact on used vehicle import and new vehicle sales. *Transp. Res. Interdiscip. Perspect.* **10**, 100384 (2021).
25. Newman, S. J., Schulte, K., Morellini, M. M., Rahal, C. & Leasure, D. R. Supplementary materials—imputation models. *Figshare* https://doi.org/10.6084/m9.figshare.23702571 (2024).

https://doi.org/10.1038/s41558-024-01943-1

Methods

Data were obtained from the Department for Transportation—a department of the government of the United Kingdom—for all 65 million privately registered used vehicles undergoing mandatory annual vehicle inspections. Traditionally termed 'MOT' tests, they were undertaken across the United Kingdom between 1 January 2005 and 31 December 2021. Used vehicles were defined as all vehicles that had undergone at least one prior inspection. These annual vehicle inspections are required by law in the United Kingdom to assess roadworthiness. They begin one year after the vehicle is first registered for motorcycles and scooters and three years after first registration for all other vehicles (in Northern Ireland, the equivalent requirement applies after four years). Alongside these data, we obtained linked vehicle-specific data on all used vehicles that had been scrapped or issued a certificate of destruction (N = 9,077,804) and all vehicles that had been flagged as exported (N = 6,922,292). Both export and scrappage certifications are supplied with exact dates.

Used vehicle summary statistics for predicted emissions, age, observed engine capacity and all other vehicle properties were calculated at a daily resolution from 1 January 2005 to 31 December 2021 inclusive, for a comprehensive sample of every on-road vehicle test (267.5 million tests; one random test per annum was selected for each vehicle) and for every scrapped (issued a vehicle scrappage certificate or certificate of destruction) or exported vehicle (Supplementary Code). We restricted our 'on-road' data to one randomly sampled roadworthy test per vehicle per year to avoid the oversampling of mechanically unreliable vehicles, which are re-tested each time they fail a roadworthy certificate (Supplementary Code). This resulted in predictions of emissions intensity across 261.1 million vehicle days, or 42,000 vehicles per day, for cars (or class 4 vehicles) known to be on road at the time of their roadworthy inspection.

All analyses were performed using R version 4.0.5 (Supplementary Code). Some 993 vehicles (0.015%) that were erroneously flagged as both exported and scrapped or destroyed were removed from analysis. As the date of first use and the date of each vehicle inspection, scrappage, destruction or export were reported, we calculated vehicle ages at each of these events. Some 60,642 (0.4%) of the reported dates were impossible and were excluded, as they were in potentially reverse order or contained typographic errors. Another 7,380 vehicles over the age of 110 years (0.05%) were excluded from analysis as they largely—but not entirely—constituted age-related coding errors.

The vehicle inspection and scrappage data were matched to a public dataset[11] of fuel efficiency and emissions data from the Vehicle Certification Agency, for 70,000 measurements tabulated by vehicle make, model, fuel type and year, measured across vehicles sold in the United Kingdom from 2000 onwards. Emissions data for carbon dioxide, carbon monoxide, nitrogen oxides, total hydrocarbon particulates and fuel efficiency of UK vehicles were also obtained from the Vehicle Certification Agency[11].Given the near-complete lack of emissions testing data for motorcycles and other vehicle classes, emissions testing data were only captured for cars and vans and were curated, quality-controlled and matched to the Department for Transport data (Supplementary Code).

Matching these emissions testing data resulted in 1.3 million exactly matched CO_2 measurements for cars in the exported or scrapped fleet (8.4% of all exported cars and 8.1% of all scrapped cars) but only 3,222 exact-matched CO_2 measurements for LGVs (0.3% of all LGVs), a number largely restricted by the abundance of rarer or untested models and the difficulty in exact matching heterogeneous make and model descriptions provided by the Vehicle Certification Agency. Exported vehicles with exact matches for measured emissions were lower (just 678 or 0.16% of all exported LGVs), excluding the reliable imputation of emissions from LGVs.

These data highlighted the extensive need for better, more comprehensive emissions testing for both new, on-road and exported vehicles. The regulatory environment needs to be restructured to fill these gaps. However, while testing regimes and data were insufficient to impute on-road emissions in motorcycle or LGV fleets, accurate imputation of car emissions and pollution was possible under the assumption that new car testing data would remain rank conserved over time. This assumption generates a lowest-possible estimate of emissions rates, under the assumption that used vehicles are 'functioning-as-new' at the point of scrappage, export or testing. This overly optimistic assumption is a key limitation of the study, one that highlights the need for far better measurement and testing of real-world emissions across all vehicle fleets.

Imputation models were constructed to capture the emissions of all exported (N = 6,072,730), scrapped (N = 9,077,804) or on-road (N = 261.1 million tests) class 4 vehicles (cars and vans below 3 T) that passed quality controls. We used the reported model year, fuel type (for example, 'Petrol/Gasoline', 'Diesel-Electric') and engine capacity (in cubic centimetres) from the annual vehicle inspection data to construct a model for each pollution type and vehicle property. Pollution types and vehicle properties that were imputed include CO_2 (g km^{-1}), total nitrous oxides (NO_x; mg km^{-1}), total particulate hydrocarbons (mg km^{-1}), carbon monoxide (mg km^{-1}) and fuel efficiency in 'miles per gallon' (MPG). Imputation models were kept deliberately simple, as we predict values across the broadest possible range of vehicles using variables that were reported at almost every vehicle inspection. This avoided overfitting and, as rare vehicles share engines and emissions technology with common makes and models, achieved accurate (Supplementary Materials and Extended Data Fig. 1) imputation for a very large (>99%) fraction of cases.

We developed imputation models using recursively partitioned regression trees[26,27], a foundational and interpretable machine learning heuristic suited to discrete effects and small variable sets (Extended Data Fig. 1). This was implemented using the 'rpart' package[26]. Models were trained on three input variables—engine size in cubic centimetres, the year of vehicle manufacture and fuel type—using tenfold random-sample cross validation, with the 'cp' model complexity parameter set to 0.001 and a minimum of 100 vehicle make models used for a parent node (that is, the 'minbucket' parameter) and a minimum of 25 vehicle make models used for a child or leaf nodes (the 'minsplit' parameter; Supplementary Code).

These models achieved high accuracy, approaching the test–retest accuracy of the Vehicle Certification Agency testing regime (Extended Data Fig. 1). For example, consecutive tests of the same vehicle for CO_2 were correlated by $r = 0.9$, whereas our recursively partitioned regression model attained an accuracy of $r = 0.94$ when imputing CO_2 for a random holdout sample (random 20% holdout sample; N = 6,708 unique vehicle makes and models; Extended Data Fig. 1a,b). We found that grid searching to further optimize model fit was therefore unnecessary, as our initial model parameters generated models that approached the highest achievable accuracy[11].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data on the exported, scrapped and used vehicle fleets supporting this study are available from the driver and vehicle standards agency (DVSA) and Department for Transport and from the authors upon explicit approval from the DVSA. Restrictions apply to these data as they contain potentially re-identifiable data on every current and former driver in Great Britain. Data on emissions testing are freely available from the Vehicle Certification Agency[13] on request from the corresponding author and in the Figshare repository[25]. Imputation models are available from the Figshare repository[25] and on request from the corresponding author.

Code availability

Imputation models and code are freely available under a Creative Commons CC-BY-NC 4.0 licence from Figshare[25] or on request from the corresponding author. Specifically, all resulting recursively partitioned regression models, code, and imputation and summary data are provided in the Supplementary Information and in open, stable repositories[25].

References

26. Therneau, T., Atkinson, B. & Ripley, B. rpart: Recursive partitioning. R version 4.1.19 http://cran.r-project.org/package=rpart (2013).
27. Zeileis, A., Hothorn, T. & Hornik, K. Model-based recursive partitioning. *J. Comput. Graph. Stat*. https://doi.org/10.1198/106186008X319331 (2008).

Acknowledgements

We would like to thank S. Sharples for her support as chief scientific officer at the Department for Transport, climate change engineer R. Jibrin for his wonderful feedback, G. Thunder for his excellent help and advice and the Driver and Vehicle Standards Agency and the Department for Transport for their substantial help and support, without which this project would not be possible. Financial Support was also gratefully received from the Leverhulme Trust for the Leverhulme Centre for Demographic Science (grant RC-2018-003; all authors), Nuffield College (C.R., M.M.M., D.R.L.), the John Fell Fund (S.J.N.; project CYD00210) and University College Oxford (S.J.N.).

Author contributions

S.J.N. designed and performed the analysis, co-designed the study, wrote the code and co-wrote the paper. K.S. performed preliminary data cleaning, developed the narrative structure and co-wrote the paper. M.M.M. contributed to the project concept, the development of the narrative structure and the drafting and revision of the paper. C.R. undertook code reviews, contributed to the drafting of the initial paper and contributed substantially to its revisions. D.R.L. contributed to the project concept, design of the study and writing the paper.

Inclusion and Ethics statement

This project was approved by the University of Oxford's Departmental Research Ethics Committee (Sociology) under ethics approval SOC_R2_001_C1A_21_66. Research data included all vehicles registered in the United Kingdom since 2005 with no information about the vehicle owners. A Research Participant Information Sheet is available from authors upon request to provide vehicle owners whose data may have been included with information about how these data were used and how to opt out of future research.

Competing interests

The authors declare no competing interests.

Additional information

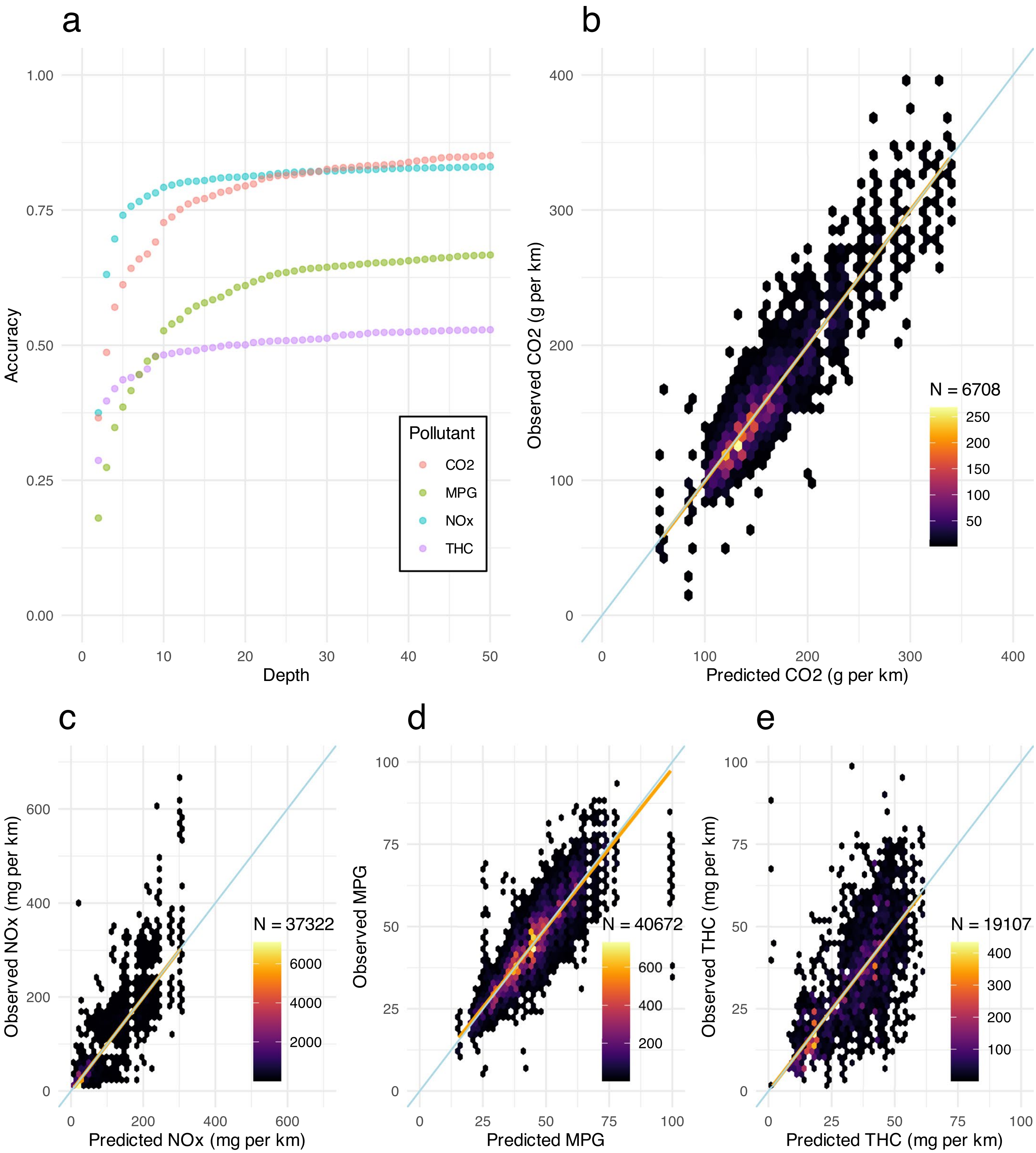
Extended data is available for this paper at https://doi.org/10.1038/s41558-024-01943-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41558-024-01943-1.

Correspondence and requests for materials should be addressed to Saul Justin Newman.

Peer review information *Nature Climate Change* thanks Pervin Kaplan, Huan Liu and Paul Wolfram for their contribution to the peer review of this work.

 https://doi.org/10.1038/s41558-024-01943-1



Extended Data Fig. 1 | Imputation accuracy for randomly sampled emissions data. Emissions of non-electric vehicles were accurately imputed from recursively partitioned regression models (Supplementary Code). Predictions made on a masked random 20% sample of vehicle emissions tests (N = 6,708 holdout make – model – year - engine combinations) showed a high degree of accuracy under ten-fold cross-validation (**a**; R^2 on y-axis). Accuracy increased with tree depth to achieve moderate to high accuracy across CO2 (**b**), nitrogen oxides (NOx; **c**), and miles per gallon (MPG; **d**) prediction models, with the lowest cross-validation accuracy for total hydrocarbon emissions (THC; **e**). Model and manufacturer effects accounted for minimal variance, independent of these effects, and were not fit to allow imputation of rare makes and models. Filled circles in (**a**) are red for CO2, green for MPG, blue for NOx, and mauve for THC.

nature portfolio

Corresponding author(s): Saul J Newman

Last updated by author(s): 22 Jan 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All data is available on request and approval from the UK Department for Transportation
Data analysis	All code is included and available from an open shared repository (link in the manuscript)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data availability statement is provided in the manuscript

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.) Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	*Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data exclusions	*Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Replication	*Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.*
Randomization	*Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.*
Blinding	*Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Observational study of vehicle emissions
Research sample	Comprehensive sample of all privately registered GB vehicles 2005-2021 inclusive
Sampling strategy	Comprehensive

Data collection	Data were obtained form the UK Department for Transportation
Timing	1 Jan 2005 - 31 Dec 2021
Data exclusions	Several quality controls described in the manuscript, exclusion criteria were not pre-established, 69015 vehicles were screened by quality controls
Non-participation	NA - participants (vehicles) were comprehensively sampled
Randomization	NA

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☒ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.

Validation

Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.

Authentication

Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.

Mycoplasma contamination

Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Palaeontology and Archaeology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals

For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Reporting on sex

Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall

	numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies
All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session
(e.g. UCSC)

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

Instrument

Identify the instrument used for data collection, specifying make and model number.

Software

Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.

Cell population abundance | *Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*

Gating strategy | *Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type | *Indicate task or resting state; event-related or block design.*

Design specifications | *Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*

Behavioral performance measures | *State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s) | *Specify: functional, structural, diffusion, perfusion.*

Field strength | *Specify in Tesla*

Sequence & imaging parameters | *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition | *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software | *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization | *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template | *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal | *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring | *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings | *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested | *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016) | *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction | *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*
Graph analysis	*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*
Multivariate modeling and predictive analysis	*Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*