Quantifying the cost savings of global solar photovoltaic supply chains

https://doi.org/10.1038/s41586-022-05316-6

Received: 30 June 2021

Accepted: 2 September 2022

Published online: 26 October 2022

Check for updates

John Paul Helveston[1], Gang He[2]✉ & Michael R. Davidson[3,4]

Achieving carbon neutrality requires deploying renewable energy at unprecedented speed and scale[1,2], yet countries sometimes implement policies that increase costs by restricting the free flow of capital, talent and innovation in favour of localizing benefits such as economic growth, employment and trade surpluses[3,4]. Here we assess the cost savings from a globalized solar photovoltaic (PV) module supply chain. We develop a two-factor learning model using historical capacity, component and input material price data of solar PV deployment in the United States, Germany and China. We estimate that the globalized PV module market has saved PV installers US$24 (19–31) billion in the United States, US$7 (5–9) billion in Germany and US$36 (26–45) billion in China from 2008 to 2020 compared with a counterfactual scenario in which domestic manufacturers supply an increasing proportion of installed capacities over a ten-year period. Projecting the same scenario forwards from 2020 results in estimated solar module prices that are approximately 20–25 per cent higher in 2030 compared with a future with globalized supply chains. International climate policy benefits from a globalized low-carbon value chain[4], and these results point to the need for complementary policies to mitigate welfare distribution effects and potential impacts on technological crowding out.

Solar energy is promised to play a crucial role in achieving a sustainable, low-carbon energy future and avoiding the worst impacts of climate change[1]. Over the past 40 years, solar photovoltaic (PV) prices have fallen by over two orders of magnitude, and during the period 2010 to 2021, the global weighted-average levelized cost of energy of newly commissioned utility-scale solar PVs fell by 88% (ref. [5]), making solar PVs cheaper than fossil fuel power in some parts of the world. Installed costs (excluding the cost of capital) fell by 81% over this period[5]. Although these dramatic price declines have been a boon for accelerating low-carbon energy deployment[6], further declines will be necessary to deploy renewables at the speed and scale that is needed to achieve climate targets, especially in the remaining parts of the world where fossil fuel power is still cheaper[7]. Recent research suggests that the rates of solar and wind energy deployment in even the fastest-deploying nations are not high enough to meet the targets necessary to avoid the worst consequences of climate change[8].

Nonetheless, rapid price declines in solar PV have not been without controversy. China, for example, has played an outsized role in scaling up the mass production of solar PV cells and modules, comprising 78% of global production in 2021[9,10] (Fig. 1). Greg Nemet went as far as to call this outcome China's "gift to the world"[11], referring to the dramatic manufacturing cost reductions achieved by Chinese firms in the past decade[5]. Yet other nations view the concentration of PV manufacturing in China as a competitive threat, and some have attributed this outcome to unfair trade practices and industrial policies implemented by China's government[12]. Countries seeking to capitalize on the growing clean energy sector are looking to protect and grow domestic manufacturers[3].

In response to these concerns, the United States and the European Union have imposed steep solar tariffs on imports from China and other countries. In June 2022, the Biden administration invoked the Defense Production Act to accelerate the onshoring of solar PV manufacturing[13]. These efforts could lead to less efficient national learning processes replacing the learning processes associated with global supply chains that have led to drastic price declines[4]. The free flow of capital (for example, foreign finance-backed start-ups), talent (for example, international collaborations with Chinese researchers) and innovations (for example, technologies pioneered in labs overseas and licensed and mass-produced in China) were essential to the rise of China's competitive solar PV industry[14]. Each of these activities is increasingly under scrutiny by the United States and other governments[15]. In the event of strict nationalization policies (including, inter alia, trade barriers in final or intermediate solar goods, restrictions on cross-national research and development, and barriers to cross-border investment), subsequent cost and performance improvements could derive primarily from activities, knowledge and capital within national borders, potentially slowing the rate of price declines in globally traded solar PV components and, consequently, the rate of solar PV deployment.

International climate policy and renewable energy deployment policy now face a crossroads: continue relying on global supply chains, or pivot towards domestic technology development and production.

[1]Department of Engineering Management and Systems Engineering, George Washington University, Washington DC, USA. [2]Department of Technology and Society, College of Engineering and Applied Sciences, Stony Brook University, Stony Brook, NY, USA. [3]Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA, USA. [4]School of Global Policy and Strategy, University of California San Diego, La Jolla, CA, USA. ✉e-mail: gang.he@stonybrook.edu

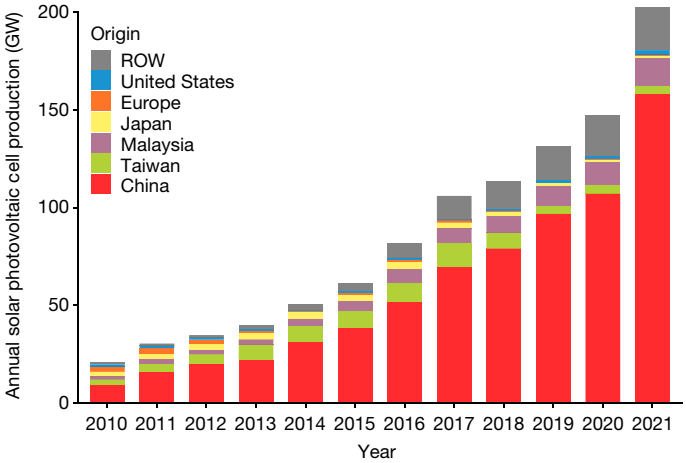


Fig. 1 | Annual solar PV cell production by origin, 2010–2021. Over the past decade, solar PV cell and module production has increasingly been concentrated in China[6]. ROW, rest of world. Data taken from ref. [9].

Table 1 | Solar PV 2030 installation targets for projection scenarios

Country	National trends		Sustainable development	
	2030 target (GW)	Implied CAGR (%)	2030 target (GW)	Implied CAGR (%)
United States	295	12	628	21
China	750	12	1,106	17
Germany	103	7	147	11
World	2,115	11	3,125	16

This study attempts to quantify the difference between these two paths in terms of the costs of deploying solar PV to achieve ambitious low-carbon goals. We collect detailed historical capacity, component and input material cost data of solar PV deployment in the United States, Germany and China, and develop a two-factor learning model to estimate a learning curve associated with the historical (globalized) solar PV supply chain. We then use these learning models to compare counterfactual historical prices and potential future prices of solar PV modules under 'global' versus 'national' market conditions. The global market scenarios reflect learning under historical market conditions whereas the national market scenarios reflect a gradual transition to fully domestically supplied markets over a ten-year period in each country.

We focus our scope on PV modules for two reasons. First, modules are a globally traded component and comprise between 20% and 40% of the installed system cost for most PV installations[16]; combined with inverters, modules comprised 61% of the global weighted-average total installed price decline between 2010 and 2020[5] (although they are expected to account for lower portions of cost in the future). Second, other 'soft costs' (such as permitting, installation and marketing) vary widely by country and have geographically limited learning and spillover effects[17]; as a result, we expect these cost components to remain relatively similar regardless of where modules are manufactured. Our analysis is limited to installed prices, not the levelized cost of energy as reflected in power purchase prices for solar energy, which also vary by country and project according to the cost of capital and other factors.

Modelling historical prices and savings

Using nation-specific, component-level price data and global PV installation and silicon price data, we estimate learning rates for solar PV modules in the three largest solar-deploying countries (China, Germany and the United States) between 2006 and 2020 using a two-factor learning model. Combined, these three markets comprised 54% of all global installed PV capacity during this period[18]. Estimated learning rates during this period are 20% in Germany, 26% in the United States and 33% in China. We then compute the counterfactual 'national markets' scenario by assuming that starting in 2006 countries began implementing nationalistic policies that gradually restrict learning to installations within their country borders over a ten-year period (for China, the starting year is 2007 owing to data availability). Annual installed capacities are assumed unchanged in the counterfactual 'national markets' scenario to provide the most policy-relevant results (see more discussion in 'Limitations' in Methods). Figure 2 shows the resulting price curves between the 'global market' and 'national market' scenarios in each country as well as the true historical prices.

Comparing the two scenarios, if each country had pursued a gradual transition to strict nationalistic policies while installing at the same rate over a ten-year period, our results imply that solar PV module prices in 2020 would have been substantially higher than their actual historical prices: 54% higher in China (US$387 per kW versus US$250 per kW), 83% in higher Germany (US$652 per kW versus US$357 per kW) and 107% higher in the United States (US$877 per kW versus US$424

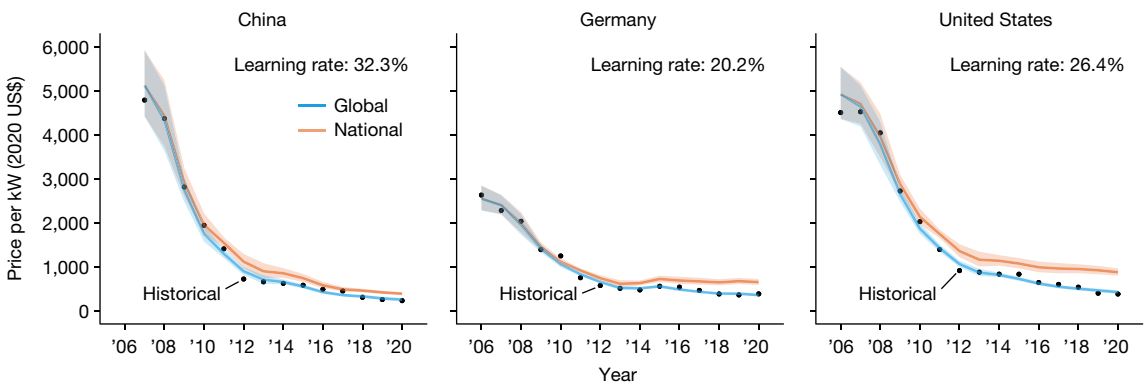


Fig. 2 | Comparison of estimated solar PV module prices under global versus national market scenarios in China (2007–2020), and Germany and the United States (2006–2020). Points are historical module prices, and the two solid lines reflect the modelled prices using global (blue) versus national (orange) markets scenarios. In each modelled curve, the learning rates are held constant by country and silicon prices follow historical global trends (Extended Data Fig. 6). The global market scenario uses global capacities and the national market scenario uses a weighted sum of national and global capacities that reflects a gradual transition to fully domestically supplied markets over a ten-year period. Uncertainty bands represent 95% confidence intervals from the estimated learning models, which were computed via simulation.

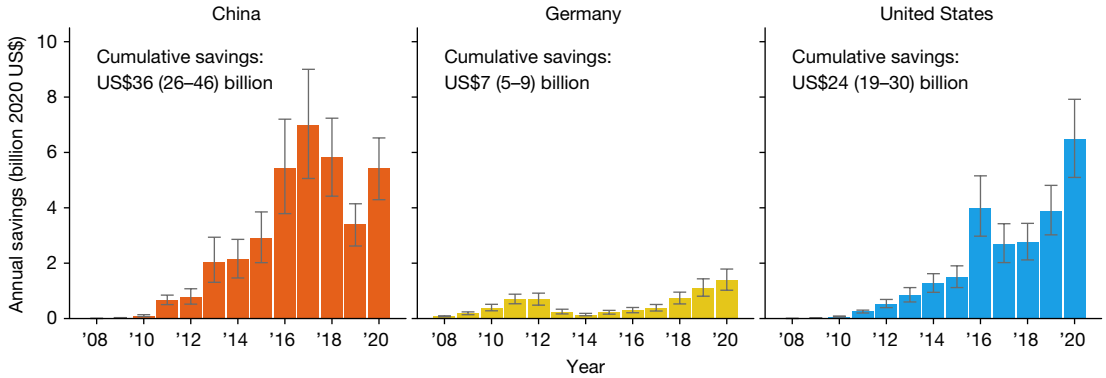


Fig. 3 | Estimated annual savings from deployed annual solar PV modules using global versus national market scenarios in China, Germany and the United States (2008–2020). Savings are calculated by multiplying the installed national capacity in each year with the difference between the modelled prices from the national and global markets scenarios. Error bars represent 95% confidence intervals computed via simulation.

per kW). Early learning, boosted in part by Germany's generous solar feed-in-tariffs, led to compounded improvements over time for the United States and China, which led to steep increases in installations in the second half of the period. The combined estimated cumulative savings across all three countries during this period from global versus national markets is US$67 billion (2020 $US), with a 95% confidence interval of US$50–84 billion (Fig. 3).

Future trajectories

As more countries introduce policies aimed at protecting local manufacturers, such as import tariffs on PV modules, continued learning-based reductions in module prices may be delayed. To assess this effect, we project solar PV module prices out to 2030 based on continued global versus national market scenarios starting from historical 2020 PV prices. These projections assume that capacity grows at a constant annual growth rate (CAGR) from 2020 installed capacity levels out to 2030 targets for each country. We consider two different future scenarios: 'national trends', which projects recent deployment trends out to 2030, and 'sustainable development', which reflects more aggressive installation growth to meet climate targets based on the Sustainable Development Scenario in the International Energy Agency *World Energy Outlook 2020*[7]. Table 1 summarizes the specific 2030 targets for each country in each scenario, and Fig. 4 shows the results of these projections.

These projections imply that prices would be substantially higher in 2030 if strict nationalistic policies were gradually implemented in each country from 2020 to 2030. Under the national trends scenario, 2030 prices would be approximately 20% higher in each country: US$162 per kW versus US$135 per kW in China, US$298 per kW versus US$251 per kW in Germany, and US$320 per kW versus US$262 per kW in the United States. Under the sustainable development scenario, the differences in prices would be approximately 25% higher in each country: US$136 per kW versus US$108 per kW in China, US$276.2 per kW versus US$220.9 per kW in Germany, and US$276.2 per kW versus US$221.3 per kW in the United States. For comparison, the US National Renewable Energy Laboratory *2021 Annual Technology Baseline* report predicts that solar PV modules will reach US$170 per kW, US$190 per kW and US$320 per kW by 2030 in advanced, moderate and conservative improvement scenarios, respectively[19]. Therefore, the differences attributed just to domestic production are up to half of the gap between worst-case and baseline cost scenarios. On the basis of the projected installed capacities, the estimated cumulative future savings from 2020 to 2030 across all three countries from global versus national markets is US$15 billion (2020 $US) with a 95% confidence interval of US$13–16 billion under the national trends scenario, and US$36 (33–39) billion under the sustainable development scenario (Extended Data Fig. 1).

Discussion

The manufacturing of solar PV modules—a globally traded commodity that is crucial to addressing climate change—is increasingly contested by governments seeking to localize benefits of the current and future scale of the industry. Yet achieving the rapid rates of solar PV deployment required to address climate change will necessarily require continued price declines at the same or greater rates as those experienced during the past decade, a period during which the free flow of global talent, capital and innovations were instrumental to cost reductions. In this paper, we contribute to understanding the implications of strict nationalistic policies by assembling component-specific solar PV price data (Extended Data Figs. 2–4) across major markets, establishing national-level estimates of learning rates that incorporate silicon prices (Extended Data Fig. 6), and quantifying the potential impact of restricted national learning on historical and projected prices and savings from solar PV deployment. The results may extend to other low-carbon technology sectors, such as wind-generating systems and electric vehicles, with caveats related to the supply chain integration and complexity of technological components. Wind-generating systems, for example, have a very globally integrated and specialized trade in intermediate components[20]; as a result, achieving 'national markets' for the entire wind supply chain could lead to even larger disruptions in terms of costs and delayed learning.

We identify three dilemmas facing policymakers in preserving established globalized supply chains: trade disputes and domestic employment, 'crowding out' of alternative technology pathways, and additional benefits and drivers of domestic sourcing. Resolving these through complementary policies that mitigate impacts on global learning are difficult but important tasks moving forwards.

Trade disputes and domestic employment

Some have attributed the concentration of PV manufacturing in China to unfair trade practices and industrial policies implemented by China's government[12]. Although constant cost multipliers would be absorbed in the national learning rates, we do not attempt to disaggregate the contributions to these rates nor do we account for changes in national-level producer subsidies or tariffs faced by importers. The 'learning curve' is a synthetic indicator that captures the cumulative effect of impacting factors on the cost evolution of a technology. Data limitations of time-varying government subsidies, industrial policies, tariffs and firm relocations prevent us from disaggregating these precise effects on price and are beyond the scope of this study.

The loss of potential manufacturing jobs in importing countries coupled with trade disputes is prompting much of the impetus for nationalistic policies. The National Renewable Energy Laboratory

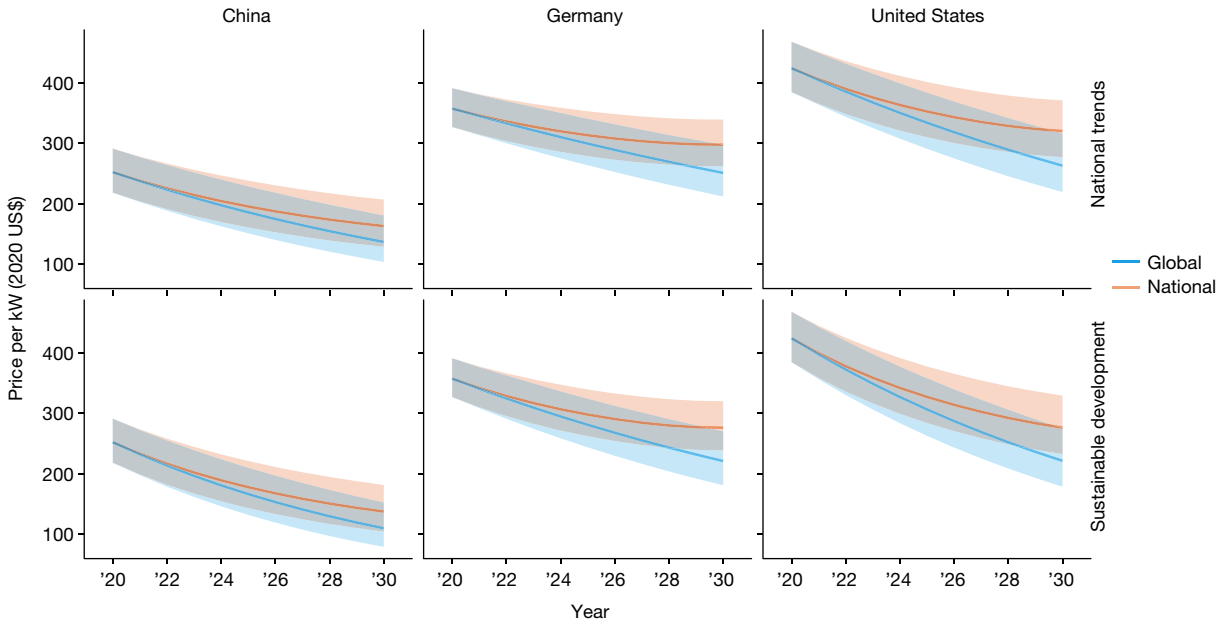


Fig. 4 | Comparison of projected solar PV module prices (2020–2030) using global versus national market scenarios in China, Germany and the United States. Projections assume CAGRs in PV installations to achieve national and global 2030 installation targets. Each curve starts at historical 2020 module prices and follows a nation-specific learning rate. In the global market scenarios, global projected installed capacities are used to project prices whereas in the national market scenarios a weighted sum of national and global capacities is used that reflects a gradual transition to fully domestically supplied markets over a ten-year period. Uncertainty bands represent 95% confidence intervals from the estimated learning models, which were computed via simulation.

estimates that there are ten times more annual jobs in system installation compared with those in the entire manufacturing supply chain (although within manufacturing, solar module production is the most labour intensive per gigawatt)[21]. Hence, if higher prices associated with nationalistic policies result in less deployment, total employment may decline, although there may be other redistributive concerns and political realities shaping preferences for certain types of jobs[22].

Our national markets counterfactual scenario is an illustrative example of more extreme decoupling, although because of the difficulty of onshoring, countries may instead opt to 'near-shore' production to a subset of countries or onshore only select parts of the supply chain. Even the three countries studied could not costlessly onshore entire supply chains; hence, our results probably represent an underestimate of the future costs of strict onshoring policies. Reciprocity in trade policies is another barrier limiting the extent to which nations can fulfil onshoring policy goals: for example, the US polysilicon industry was once a dominant global supplier to solar PV manufacturers but became the first casualty of the solar trade war between China and the United States when China retaliated for tariffs on imported Chinese modules.

Technological 'crowding out'

Some have argued that the rapid price declines of monocrystalline silicon (c-Si) PV cells, driven in part by Chinese industrial policies to ramp up production in China, might have 'crowded out' other emerging solar technologies, such as 'thin film' solar cells for which the United States has a sizable global market share and that could have achieved even lower prices without fierce competition from c-Si[23,24]. Such an argument is not without precedent. For example, ref. [25] found that offshoring manufacturing in the optoelectronics industry to developing East Asia led to such notable price reductions in the incumbent technology that emerging and potentially groundbreaking technologies could not compete and were largely abandoned.

Although these concerns are not without merit, they are not necessarily the only forces at play in the global PV industry. Indeed, PV cell and module manufacturing has followed a developmental path common to many industries in which initial, intense experimentation is followed by the emergence of a 'dominant design'[26] and a shift in productive activity away from product innovations and towards production improvements to increase scale and reduce costs[27–30]. This shift in focus towards production tends to precipitate two related phenomena: (1) unit costs drop dramatically as firms identify successful production innovations, and (2) many competing firms fail as production tends to concentrate around the handful of firms that are able to compete on lower costs. In some industries, this also coincides with offshoring production in search of lower-cost production environments, although this is not always the case[3]. Thus, it remains unclear whether the concentration of PV cell and module production in China was purely a result of government intervention or perhaps a combination of factors, such as the natural evolution of a maturing industry[31]. Chinese policies may have accelerated cost declines in c-Si cells and modules, but whether they alone led to the crowding out of other potential technologies remains debatable.

Additional domestic and diversified sourcing drivers

A domestic manufacturing base in solar PV may provide other benefits besides direct employment worthy of future study. Our model does not incorporate any spillover benefits to adjacent industries, such as semiconductors and electronics. For example, polysilicon production is part of both advanced chip and solar supply chains, although solar-grade polysilicon has purity requirements several orders of magnitude lower[32]. Establishing a stronger link between public funding of research and development and the private sector has been identified as important to achieving climate technology innovation goals, both by reducing the risks of scale-up and by providing access to markets[33]. Foreign manufacturers may be undesirable or infeasible partners with public money. However, private sector-led efforts can be effective internationally: Chinese solar firms largely innovated through improved manufacturing processes and strategic international partnerships, including with US-based start-ups unable to scale domestically[34].

Reliance on a single or small set of countries in crucial supply chain bottlenecks, even if reducing costs and enhancing learning, may generate risks of disruption based on natural disasters or geopolitical conflict. A managed diversification—instead of national onshoring—could provide a pathway to mitigate the cost impacts of hardening supply chains.

Finally, maintaining adequate environmental, health and labour standards in the production of traded goods is important for ethical reasons and is increasingly raised in the context of maintaining a level-playing field in trade agreements. The Xinjiang region of China, where much of the world's solar-grade polysilicon is produced, has come under increased scrutiny owing to allegations of forced labour. The solar industry has responded with proposed traceability protocols, which if effective could obviate the need to onshore production for ethical reasons[35]. Further work is needed on the feasibility of such protocols.

This study presents a quantitative estimation of the historical and future cost savings from a globalized solar PV supply chain. The results provide evidence of the benefits of global learning processes in terms of achieving lower prices to accelerate low-carbon technology deployment, which could potentially be delayed by emerging nationalistic policy efforts. When negotiators meet to discuss accelerating action towards the goals of the Paris Agreement, and when policymakers plan for pathways to achieve mid-century carbon neutrality, they should recognize that these aspirations may be difficult or impossible to achieve without globalized low-carbon supply chains. Complementary policies are necessary to address dilemmas and debates with respect to localizing manufacturing and to ensure continued price declines.

Online content

Any methods, additional references, Nature Research 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/s41586-022-05316-6.

1. *Net Zero by 2050* (IEA, 2021).
2. Doerr, J. & Panchadsaram, R. *Speed & Scale: An Action Plan for Solving Our Climate Crisis Now* (Portfolio, 2021).
3. Sarah L. et al. *Industrial Policy, Trade, And Clean Energy Supply Chains* (CSIS & BloombergNEF, 2021); https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/210224_Ladislaw_Industrial_Policy.pdf.
4. Goldthau, A. & Hughes, L. Protect global supply chains for low-carbon technologies. *Nature* **585**, 28–30 (2020).
5. *Renewable Power Generation Costs in 2021* (IRENA, 2021); https://www.irena.org/publications/2022/Jul/Renewable-Power-Generation-Costs-in-2021.
6. Helveston, J. & Nahm, J. China's key role in scaling low-carbon energy technologies. *Science* **366**, 794–796 (2019).
7. *World Energy Outlook 2020* (IEA, 2020).
8. Cherp, A., Vinichenko, V., Tosun, J., Gordon, J. A. & Jewell, J. National growth dynamics of wind and solar power compared to the growth required for global climate targets. *Nat. Energy* **6**, 742–754 (2021).
9. Jäger-Waldau, A. Snapshot of photovoltaics—February 2022. *EPJ Photovolt.* **13**, 9 (2022).
10. *Special Report on Solar PV Global Supply Chains* (IEA, 2022); https://www.iea.org/reports/solar-pv-global-supply-chains.
11. Nemet, G. F. *How Solar Energy Became Cheap: A Model for Low-Carbon Innovation* (Routledge, 2019).
12. Atkinson, R. D. Why China needs to end its economic mercantilism. *HuffPost* https://www.huffpost.com/entry/why-china-needs-to-end-it_b_84028 (2008).
13. *Fact Sheet: President Biden Takes Bold Executive Action to Spur Domestic Clean Energy Manufacturing* (The White House, 2022).
14. Green, M. A. How did solar cells get so cheap? *Joule* **3**, 631–633 (2019).
15. Tillman, B. Red scare or red herring: how the "China Initiative" strategy for non-traditional collectors is stifling innovation in the United States. *Seattle J. Technol. Environ. Innov. Law* **11**, 6 (2020).
16. Fu, R., Feldman, D. & Margolis, R. *U.S. Solar Photovoltaic System Cost Benchmark: Q1 2018* (NREL, 2018); https://www.nrel.gov/docs/fy19osti/72399.pdf.
17. Nemet, G. F., Lu, J., Rai, V. & Rao, R. Knowledge spillovers between PV installers can reduce the cost of installing solar PV. *Energy Policy* **144**, 111600 (2020).
18. *Renewable Energy Statistics 2021* (IRENA, 2021); https://www.irena.org/publications/2021/Aug/Renewable-energy-statistics-2021.
19. *2021 Annual Technology Baseline* (NREL, 2021); https://atb.nrel.gov/electricity/2021/data
20. Surana, K., Doblinger, C., Anadon, L. D. & Hultman, N. Effects of technology complexity on the emergence and evolution of wind industry manufacturing locations along global value chains. *Nat. Energy* **5**, 811–821 (2020).
21. Feldman, D. & Margolis, R. *H2 2020 Solar Industry Update* (NREL, 2021); https://www.nrel.gov/docs/fy21osti/79758.pdf.
22. Chung, D., Horowitz, K. & Kurup, P. *On the Path to SunShot: Emerging Opportunities and Challenges in U.S. Solar Manufacturing* (NREL, 2016); https://www.nrel.gov/docs/fy16osti/65788.pdf.
23. Hart, D. *The Impact of China's Production Surge on Innovation in the Global Solar Photovoltaics Industry* (ITIF, 2020); https://itif.org/publications/2020/10/05/impact-chinas-production-surge-innovation-global-solar-photovoltaics.
24. Sivaram, V., Dabiri, J. O. & Hart, D. M. The need for continued innovation in solar, wind, and energy storage. *Joule* **2**, 1639–1642 (2018).
25. Fuchs, E. & Kirchain, R. Design for location? The impact of manufacturing offshore on technology competitiveness in the optoelectronics industry. *Manage. Sci.* **56**, 2323–2349 (2010).
26. Abernathy, W. J. & Utterback, J. M., others. Patterns of industrial innovation. *Technol. Rev.* **80**, 40–47 (1978).
27. Gort, M. & Klepper, S. Time paths in the diffusion of product innovations. *Econ. J.* **92**, 630–653 (1982).
28. Utterback, J. M. & Suárez, F. F. Innovation, competition, and industry structure. *Res. Policy* **22**, 1–21 (1993).
29. Utterback, J. M. *Mastering the Dynamics of Innovation: How Companies Can Seize Opportunities in the Face of Technological Change* (Harvard Business School, 1994).
30. Agarwal, R. & Gort, M. The evolution of markets and entry, exit and survival of firms. *Rev. Econ. Stat.* **78**, 489–498 (1996).
31. Carvalho, M., Dechezleprêtre, A. & Glachant, M. *Understanding the Dynamics of Global Value Chains for Solar Photovoltaic Technologies.* Economic Research Working Paper No. 40 (WIPO, 2017).
32. *Building Resilient Supply Chains, Revitalizing American Manufacturing, and Fostering Broad-Based Growth* (The White House, 2021); https://www.whitehouse.gov/wp-content/uploads/2021/06/100-day-supply-chain-review-report.pdf.
33. Myslikova, Z. & Gallagher, K. S. Mission Innovation is mission critical. *Nat. Energy* **5**, 732–734 (2020).
34. Nahm, J. & Steinfeld, E. S. Scale-up nation: China's specialization in innovative manufacturing. *World Dev.* **54**, 288–300 (2014).
35. *Solar Supply Chain Traceability Protocol 1.0* (SIEA, 2021); https://www.seia.org/research-resources/solar-supply-chain-traceability-protocol.

Methods

Learning models and simulations

The learning curve model is widely used to describe the evolution of production costs for technologies as they scale up[36–40]. In its simplest form, the learning curve defines a relationship in log–log space between cost (or price) and cumulative capacity[41]. The model can be expanded to incorporate not only the processes of 'learning by doing' but also 'learning by researching' and changes in material input prices[42,43]. Here we adopt a two-factor learning model relating the unit price in year t and country i of solar PV modules, p_{it}, to the cumulative installed PV capacity in year t, q_t, and globally averaged polysilicon prices in year t, s_t (the primary input material to PV modules):

$$\ln p_{it} = \ln\alpha_i + \beta_i \ln q_t + \gamma_i \ln s_t \quad (1)$$

Here, α_i is a constant related to starting year conditions in country i, γ_i measures the sensitivity to polysilicon prices, and β_i is the learning coefficient in country i, which is related to the learning rate (L_i) by:

$$L_i = 1 - 2^{\beta_i} \quad (2)$$

For each country, i, we estimate learning coefficients (Extended Data Table 1), β_i, under historical 'global market' conditions using linear least-squares regression on equation (1). These learning models set a baseline for learning rates under historical market conditions and assume that variations in country-level module pricing were due to transportation, administrative and other non-learning costs.

We then construct counterfactual 'national market' scenarios by assuming that the learning-related price decreases in country i from the starting year, t_0, are derived from incrementally more nationally installed PV capacity:

$$q_t - q_{t-1} = (q_{it} - q_{it-1}) + (1 - \lambda_t)(q_{jt} - q_{jt-1}) \quad (3)$$

where q_{it} is the cumulative installed capacity in country i in year t, q_{jt} is the cumulative installed capacity in all other countries in year t, and λ_t is a value ranging from 0 to 1. This defines a scenario whereby incremental capacity installed in each year increasingly comes from national as opposed to global installations as λ_t shifts from 0 to 1. In our baseline simulations, λ_t ranges from 0.1 to 1.0 in increments of 0.1 as t goes from 1 to 10, simulating a gradual transition to a scenario where all new national PV capacity is domestically supplied. At the starting year of both the historical and projection scenarios, $\lambda_t = 0$ and the cumulative capacity is set to the globally installed capacity in that year. Unit price declines under national market conditions thus evolve more slowly according to how rapidly λ_t approaches 1.

The national market scenarios propose that national-specific learning is proportionally derived from national versus global cumulative installed capacities, and by definition $q_{it} < q_t$. Extended Data Fig. 5 illustrates the relationship between λ_{it} and the proportion of national to global cumulative installed capacity over all years for each country. It is noted that the same value of λ_i does not translate to the same proportion of national learning for each country. For example, if $\lambda_i = 0.4$, then the proportion of national learning is 15% in the United States, 44% in China and 40% in Germany.

Uncertainty in parameter estimates is propagated throughout all of our analyses using multivariate normal draws from the full covariance matrix of model parameters. Lower and upper bounds on results reflect a 95% confidence interval taken from the 2.5% and 97.5% percentiles from these draws.

Limitations

Learning-rate analyses, although widely used, are subject to critiques in terms of under-specifying learning mechanisms[40,44,45]. In our application of these models, we include exogenous factors that could influence module prices but are not directly linked to learning (for example, polysilicon prices). Otherwise, we estimate a single learning coefficient for each country that captures the average learning owing to a variety of nation-specific factors that contribute to learning, such as learning by doing (average plant size), learning by researching (research and development) and so on. Although other studies have estimated learning models that attempt to disaggregate learning into constituent components[46], our research focuses on the nation-specific price implications of trade barriers. Data gaps and insufficient observations preclude explaining the contributing factors to learning in each country. This introduces potential biases if learning mechanisms are differentially affected by globalization. Given the concentration of PV panel manufacturing in China, it is possible that a portion of the learning in China was due to achieving higher economies of scale than manufacturers in the United States and Germany. If so, then the savings reported from the differences in the global versus national market scenarios may be overestimated, assuming that US and German manufacturers would have achieved similar economies of scale in a counterfactual scenario where national producers meet domestic demand. Three alternative models were estimated to disaggregate module production, installation capacity and average plant size. Those results are shown in Extended Data Tables 2–4. Improving ease of access to credit for solar projects, as reflected in declining trends in weighted-average cost of capital, has and will continue to have a large impact on reducing power purchase prices for solar[5,47]. Therefore, restrictions in capital flows following from nationalistic policies could lead to even larger costs on developers. Installed capacities are unchanged across scenarios, ignoring any effects of price elasticity of demand which, if incorporated, would result in fewer installations in higher-cost 'national markets' scenarios. Finally, the specific outcomes in terms of estimating savings from global versus national market scenarios are sensitive to simulation parameters, such as the number of years until all national capacity is domestically supplied. These parameters can be varied and the outcomes compared using an open-source application available at https://jhelvy.shinyapps.io/solar-learning-2021/.

Reporting summary

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

Data availability

We compile a comprehensive dataset of historical solar capacity and component price globally and in the United States, China and Germany. All data are publicly available at https://doi.org/10.5281/zenodo.6989075. Global installed PV capacity and price data are from the open database of the International Renewable Energy Agency (IRENA)[18]. For the United States, solar capacity data are from the Solar Energy Industries Association (SEIA)[48], and module prices are assembled from two sources: the Lawrence Berkeley National Laboratory (LBNL)[49] and the National Renewable Energy Laboratory (NREL)[16]. The LBNL data are used for the 2006–2018 period as this series ends in 2018, and the NREL data are used for 2019–2020 to extend the series to 2020. This was chosen because the NREL data only start in 2010, and thus the LBNL series covers a broader range (Extended Data Figs. 2–4). For China, both the installed capacity and module price data (2007–2018) were extracted from reports and presentations by the Energy Research Institute (ERI)[50], and the 2019–2020 data were extracted from China Photovoltaic Industry Association where the historical data are identical to that of ERI[51]. For Germany, capacity data are from IRENA, and module price data were extracted from Fraunhofer ISE[52]. All prices are in 2020 US$, and we adopt inflation adjustments using the IMF (https://data.imf.org/) and exchange rates from the Federal Reserve Bank (https://www.federalreserve.gov/releases/h10/hist/). Source data are provided with this paper.

Code availability

All of the code used to process the data and produce all analyses and figures is publicly available at https://doi.org/10.5281/zenodo.6989075.

36. McDonald, A. & Schrattenholzer, L. Learning rates for energy technologies. *Energy Policy* **29**, 255–261 (2001).
37. Nemet, G. F. Beyond the learning curve: factors influencing cost reductions in photovoltaics. *Energy Policy* **34**, 3218–3232 (2006).
38. Qiu, Y. & Anadon, L. D. The price of wind power in China during its expansion: technology adoption, learning-by-doing, economies of scale, and manufacturing localization. *Energy Econ.* **34**, 772–785 (2012).
39. Zheng, C. & Kammen, D. M. An innovation-focused roadmap for a sustainable global photovoltaic industry. *Energy Policy* **67**, 159–169 (2014).
40. Rubin, E. S., Azevedo, I. M. L., Jaramillo, P. & Yeh, S. A review of learning rates for electricity supply technologies. *Energy Policy* **86**, 198–218 (2015).
41. Yelle, L. E. The learning curve: historical review and comprehensive survey. *Decis. Sci.* **10**, 302–328 (1979).
42. Yu, C. F., van Sark, W. G. J. H. M. & Alsema, E. A. Unraveling the photovoltaic technology learning curve by incorporation of input price changes and scale effects. *Renew. Sustain. Energy Rev.* **15**, 324–337 (2011).
43. Zhang, C., Xie, L., Qiu, Y., (Lucy) & Wang, S. Learning-by-manufacturing and learning-by-operating mechanisms drive energy conservation and emission reduction in China's coal power industry. *Resour. Conserv. Recycl.* **186**, 106532 (2022).
44. Lewis, J. I. & Nemet, G. F. Assessing learning in low carbon technologies: toward a more comprehensive approach. *WIREs Clim. Change* **12**, e730 (2021).
45. Meng, J., Way, R., Verdolini, E. & Anadon, L. D. Comparing expert elicitation and model-based probabilistic technology cost forecasts for the energy transition. *Proc. Natl Acad. Sci. USA* **118**, e1917165118 (2021).
46. Kavlak, G., McNerney, J. & Trancik, J. E. Evaluating the causes of cost reduction in photovoltaic modules. *Energy Policy* **123**, 700–710 (2018).
47. Vartiainen, E., Masson, G., Breyer, C., Moser, D. & Román Medina, E. Impact of weighted average cost of capital, capital expenditure, and other parameters on future utility-scale PV levelised cost of electricity. *Prog. Photovolt. Res. Appl.* **28**, 439–453 (2020).
48. *Solar Industry Research Data* (SEIA, 2021); https://www.seia.org/solar-industry-research-data
49. Barbose, G. L. & Darghouth, N. R. *Tracking the Sun: Pricing and Design Trends for Distributed Photovoltaic Systems in the United States—2019 Edition* (LBNL, 2019); https://emp.lbl.gov/publications/tracking-sun-pricing-and-design.
50. Wang, S. *The Status and Perspectives of China's PV Industry.* Clean Energy Summit 2019. (2019).
51. Wang, B. *PV Industry in 2020, and Perspectives for 2021.* China Photovoltaic Industry Association. (2020).
52. Wirth, H. *Recent Facts about Photovoltaics in Germany* (Fraunhofer ISE, 2021); https://www.ise.fraunhofer.de/content/dam/ise/en/documents/publications/studies/recent-facts-about-photovoltaics-in-germany.pdf.

Acknowledgements We thank G. Nemet, G. Barbose, N. Darghouth, H. Bahar, A. Jäger-Waldau and P. Mints for their help in data sharing and answering our data questions; and D. Hart and the Information Technology and Innovation Foundation for hosting the Energy Innovation and Climate-tech 'Boot Camp' for early career scholars (funded by the Alfred P. Sloan Foundation) where many of the initial conversations around this study began.

Author contributions G.H. initiated the research idea. J.P.H. led data curation. M.R.D. wrote the initial analysis code, and J.P.H. wrote the final analysis and visualization code. All authors contributed equally to conceptualization and writing.

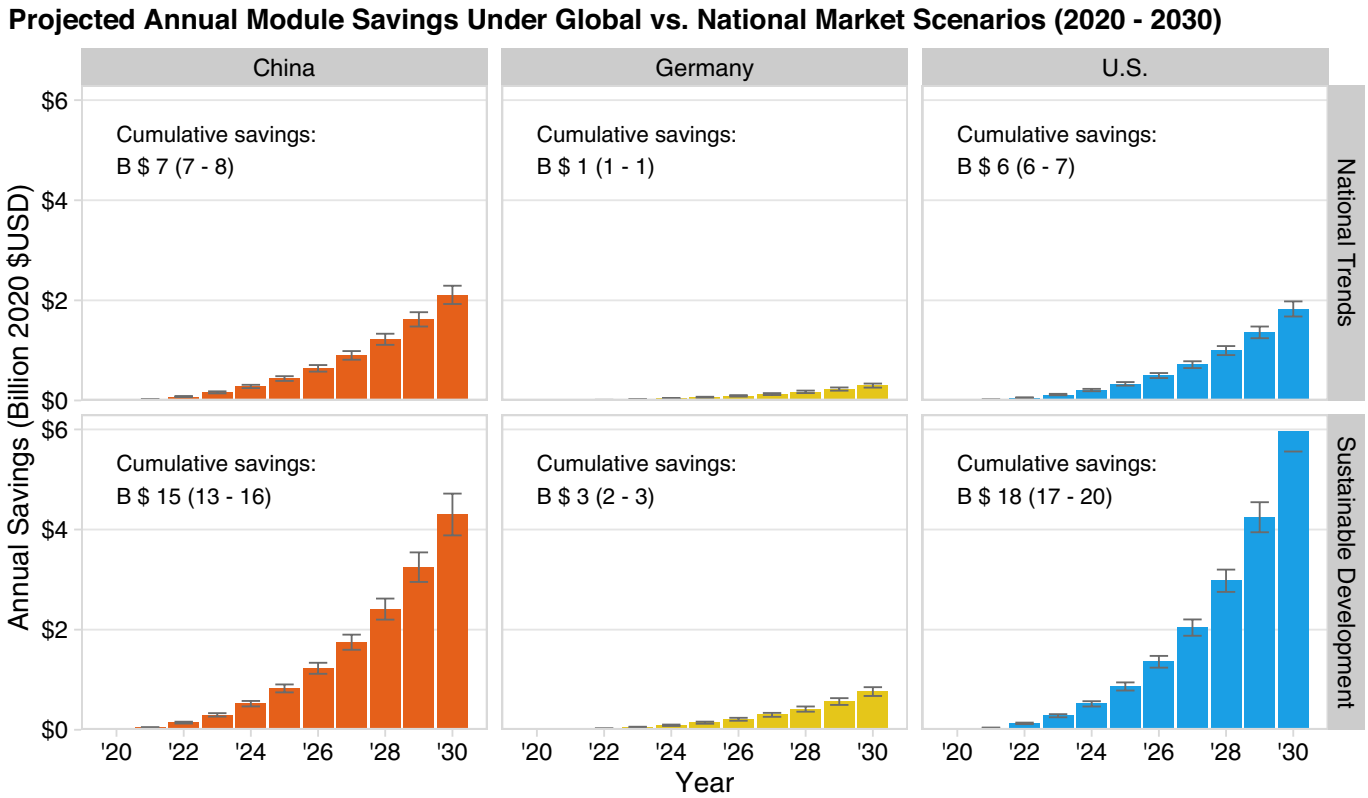
Competing interests The authors declare no competing interests.

Additional information

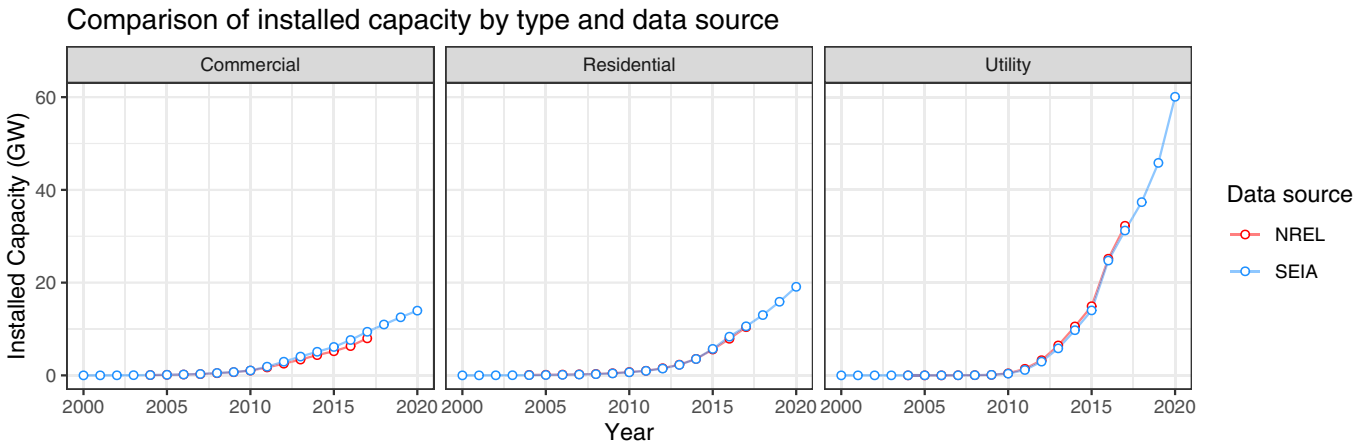
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05316-6.

Correspondence and requests for materials should be addressed to Gang He.

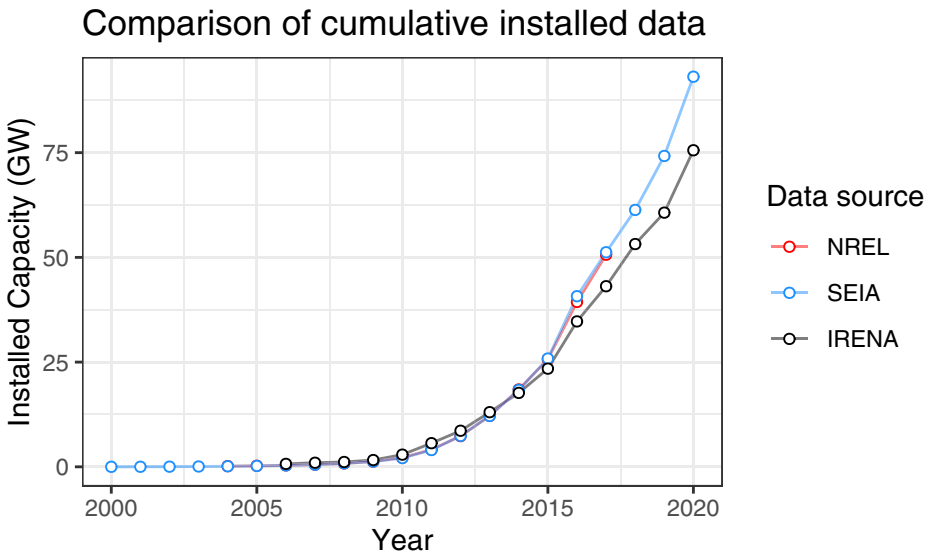
Peer review information *Nature* thanks Aleh Cherp, M. Chiesa, Paul Drummond and Yueming Qiu for their contribution to the peer review of this work. Peer reviewer reports are available.



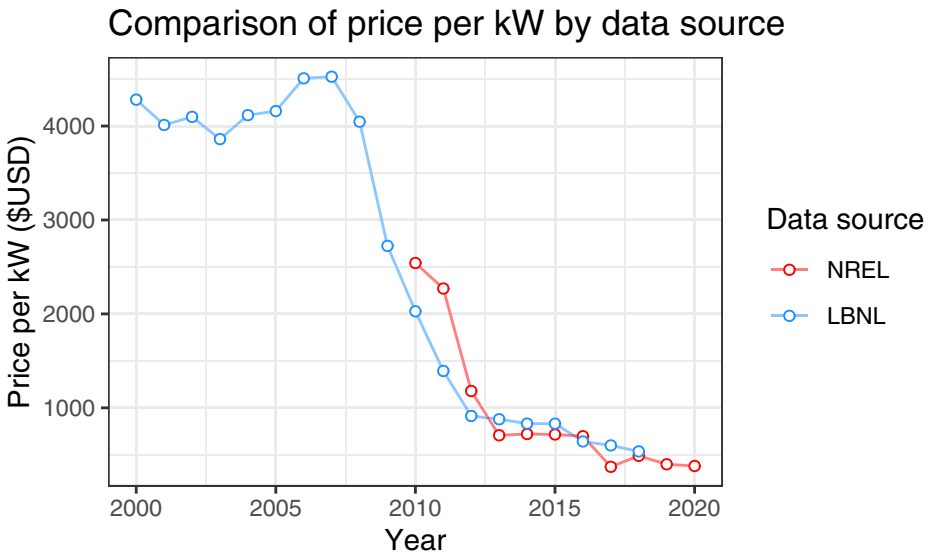
Extended Data Fig. 1 | Comparison of projected annual savings (2020–2030) using global versus national market scenarios in China, Germany and the United States. Savings are calculated by multiplying the installed national capacity in each year with the difference between the modelled prices from the national and global markets scenarios. Error bars represent 95% confidence intervals computed via simulation.



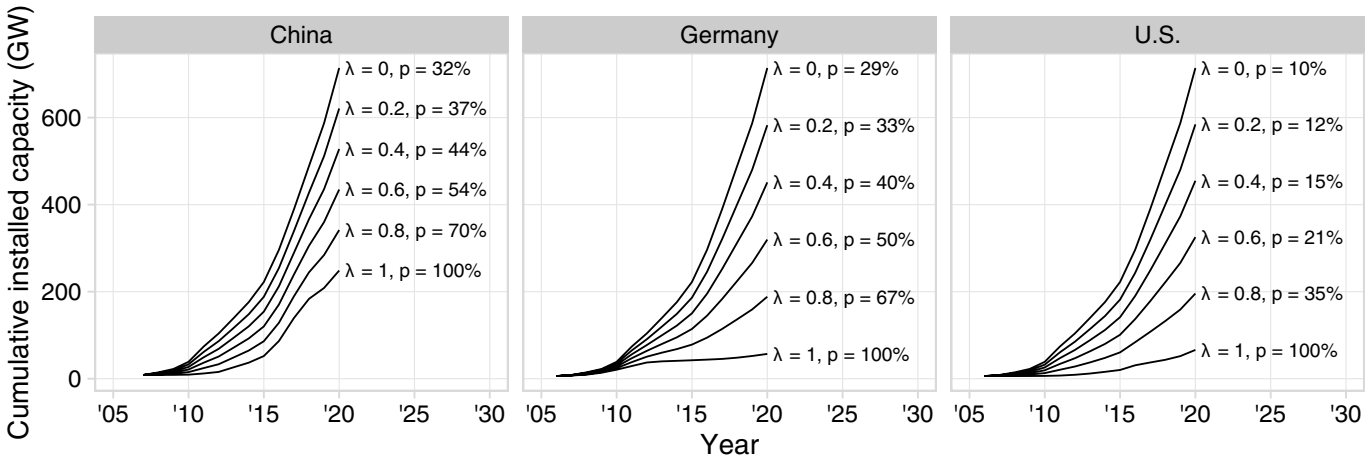
Extended Data Fig. 2 | Comparison of the US installed solar PV capacity by type and data source. The data largely agree between NREL[16] and SEIA[48]. However, SEIA data are updated to 2020 and therefore are used in this study.



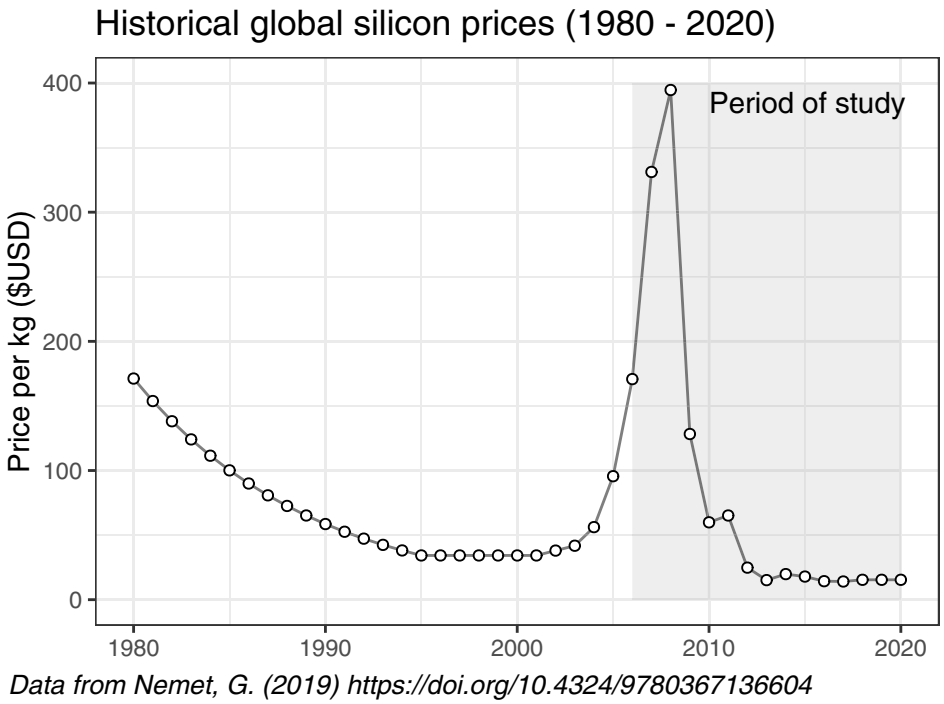
Extended Data Fig. 3 | Comparison of the US cumulative installed solar PV capacity by data source. The data largely agree between NREL[16] and SEIA[48] while the data from IRENA[18] suggest slightly lower installed capacities in the last five years.



Extended Data Fig. 4 | Comparison of the US solar PV module prices by data source. We used the LBNL data[49] for the 2006–2018 period in this study as this series ends in 2018, as well as the NREL data[16] for 2019–2020 to extend the series to 2020. We opted for this because the NREL data only start in 2010, and thus the LBNL series covers a broader range.



Extended Data Fig. 5 | Relationship between λ and the proportion of national to global cumulative installed capacity (2006–2020). The same value of λ does not translate to the same proportion of national learning for each country. For example, if $\lambda = 0.4$, then the proportion of national learning is 15% in the United States, 44% in China and 40% in Germany.



Extended Data Fig. 6 | Historical global silicon prices (1980–2020)[8]. Silicon is a key material input but is not directly linked to learning. Silicon prices experienced a major spike from US$171 per kg in 2006 to a peak at US$395 per kg in 2008, which could influence module prices notably, so we include this in our two-factor learning model.

Extended Data Table 1 | Estimated learning model coefficients

	United States Est. (Std. Err.)	China Est. (Std. Err.)	Germany Est. (Std. Err.)
(Intercept)	15 (1.04)***	18 (1.58)***	12 (0.96)***
log(cum_capacity_kw)	-0.44 (0.045)***	-0.57 (0.070)***	-0.33 (0.042)***
log(price_si)	0.15 (0.058)*	0.23 (0.079)	0.21 (0.054)

*p<0.05; **p<0.01; ***p<0.001

This table reports correlations between cumulative installed capacities (cum_capacity_kw), polysilicon prices (price_si) and unit prices of solar PV modules: the two-factor learning model that we construct in this study. Standard errors are reported in parentheses. Asterisks indicate the level of significance: *5%; **1%; ***0.1%.

Extended Data Table 2 | Estimated learning model coefficients from alternative model 1, which includes an additional covariate for cumulative national module production capacity

	United States Est. (Std. Err.)	China Est. (Std. Err.)	Germany Est. (Std. Err.)
(Intercept)	16 (2.47)***	15 (4.89)*	18 (2.28)***
log(cum_installed_kw)	-0.23 (0.347)	0.04 (0.906)	-0.14 (0.103)
log(cum_production_kw)	-0.35 (0.505)	-0.42 (0.686)	-0.52 (0.203)*
log(price_si)	0.05 (0.146)	0.21 (0.216)	0.02 (0.105)

*p<0.05; **p<0.01; ***p<0.001

Alternative model 1 generates a multi-factor log-log learning model with covariates for cumulative national module production capacity (cum_production_kw) and cumulative global installed capacity (cum_installed_kw). However, these two variables are highly correlated in each country. This table shows that adding cumulative national production capacity into the model eliminates the statistical significance of the cumulative installed capacity term for all countries, rendering the coefficients unidentifiable. In addition, the standard errors (reported in parentheses) are likely to be artificially low as standard errors are not correctly computed in basic linear regression when highly correlated variables are included in the regression. Thus, including national module production capacity directly in the regression is an infeasible approach. Asterisks indicate the level of significance: *5%; **1%; ***0.1%.

Extended Data Table 3 | Estimated learning model coefficients from alternative model 2, which includes an additional covariate for cumulative national installed capacity

	United States Est. (Std. Err.)	China Est. (Std. Err.)	Germany Est. (Std. Err.)
(Intercept)	15 (1.08)***	19 (2.09)***	15 (2.14)***
log(cum_installed_kw)	-0.41 (0.073)***	-0.60 (0.166)**	-0.26 (0.063)**
log(cum_installed_kw_i)	-0.04 (0.077)	0.03 (0.103)	-0.21 (0.157)
log(price_si)	0.16 (0.064)*	0.18 (0.117)	0.16 (0.063)*

*p<0.05; **p<0.01; ***p<0.001

Alternative model 2 considers adding cumulative national installed capacity (cum_installed_kw_i) into the model. This table shows that the nation-specific installed capacity terms are not statistically significant. Furthermore, the primary coefficient of interest, β_i, is very close to those in our main regression model (see Extended Data Table 1). Asterisks indicate the level of significance: *5%, **1%, ***0.1%.

Extended Data Table 4 | Estimated learning model coefficients from alternative model 3, which includes an additional covariate for global average plant size

	United States Est. (Std. Err.)	China Est. (Std. Err.)	Germany Est. (Std. Err.)
(Intercept)	15 (1.57)***	17 (2.25)***	13 (1.60)***
log(cum_installed_kw)	-0.32 (0.150)	-0.37 (0.216)	-0.49 (0.153)*
log(ave_plant_size_kw)	-0.22 (0.166)	-0.22 (0.239)	0.19 (0.170)
log(price_si)	0.20 (0.140)	0.38 (0.201)	0.09 (0.143)

*p<0.05; **p<0.01; ***p<0.001

Alternative model 3 controls for economies of scale (EOS) by including a proxy variable: global average plant size (ave_plant_size_kw). This table shows that including the plant size data eliminates the statistical significance of all terms. Asterisks indicate the level of significance: *5%; **1%; ***0.1%.

nature portfolio

Corresponding author(s): Gang He

Last updated by author(s): Aug 13, 2022

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	We compile a comprehensive dataset of historical solar capacity and component price globally and in the U.S., China, and Germany. All code and data are publicly available at https://doi.org/10.5281/zenodo.6989075. Global installed PV capacity and price data are from the open database of the International Renewable Energy Agency (IRENA)15 (https://www.irena.org/statistics). In the U.S, solar capacity data are from the Solar Energy Industries Association (SEIA)44, and module prices are assembled from two sources: the Lawrence Berkeley National Laboratory (LBNL)45 and the National Renewable Energy Laboratory (NREL)13. The LBNL data are used for the 2006 - 2018 period since this series ends in 2018, and the NREL data are used for 2019 - 2020 to extend the series to 2020. This was chosen because the NREL data only start in 2010, and thus the LBNL series covers a broader range (Extended Data Figures 3-4). For China, both the installed capacity and module price data (2007 - 2018) were extracted from reports and presentations by the Energy Research Institute (ERI)46, and the 2019-2020 data were extracted from China Photovoltaic Industry Association where the historical data are identical to that of ERI47. For Germany, capacity data are from IRENA, and module price data were extracted from Fraunhofer ISE48. All prices are in $2020 USD, and we adopt inflation adjustments using IMF (https://data.imf.org/) and exchange rates from the Federal Reserve Bank (https://www.federalreserve.gov/releases/h10/hist/).
Data analysis	R 4.0.3; RStudio 2022.02.1

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

All of the raw data as well as the code used to process the data and produce all analyses and figures are publicly available on Github at https://github.com/jhelvy/solar-learning-2021.

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	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). 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. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*

Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

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
☐	☐ Human research participants
☐	☐ Clinical data
☐	☐ Dual use research of concern

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

Cell line source(s)	*State the source of each cell line used.*
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 organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex 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, sex 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.*
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.

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, 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.

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

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*

Sequencing depth	*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.*