Institute for
New Economic Thinking
AT THE OXFORD MARTIN SCHOOL

Pivoting from Dirty to Clean: The Intellectual Distance between Clean and Dirty Technologies

Su Jung Jee & Sugandha Srivastav

Nov 2021

INET Oxford Working Paper No. 2021-22

Pivoting from Dirty to Clean:
The Intellectual Distance between Clean and Dirty Technologies

Su Jung Jee[1*] and Sugandha Srivastav[1,2]

Do clean technologies learn from their dirty counterparts? Using patents granted by USPTO from 1976 to 2020, we evaluate the "intellectual distance" between clean and dirty technologies. Our measure of intellectual distance is intuitively similar to "degrees of separation" where 1 indicates that a clean patent directly cites a prior dirty patent, and 2 indicates that there is an intermediary technology. We find that less than one-tenth of clean patents directly cite prior dirty art. Since citations are a proxy for learning, this implies that for the most part, leveraging dirty knowledge to pivot into clean sectors is not straightforward. However, there is a high degree of heterogeneity. Some clean technologies such as geothermal energy, carbon capture and storage, and offshore wind learn significantly from dirty technologies, due to shared knowhow related to drilling, pollution-control and operating out at sea. Our analysis identifies "clean adjacent sectors" that build upon dirty knowledge inputs, which could be plausible diversification options for dirty firms.

Keywords: Green Transition, Intellectual Distance, Clean Technology, Dirty Technology, Intellectual Carbon Lock-in

[1] *Institute for New Economic Thinking at the Oxford Martin School, University of Oxford,* [2] *Smith School of Enterprise and Environment, University of Oxford, UK *email: su.jee@maths.ox.ac.uk; sugandha.srivastav@ouce.ox.ac.uk*

Economic activity has an imperative to transition to a low-carbon paradigm to limit the impacts of climate change. However, fossil fuel firms make up a quarter of the total value of global equity markets, account for more than half of the non-financial corporate bond market and play a prominent role in some of the world's largest economies such as India, China, USA, and Russia (Jamasmie, 2020). A high degree of carbon lock-in co-exists with an imperative to transition to clean sectors. For dirty firms, the pivot towards clean industries is not only a matter of honoring climate pledges but also of mitigating transitional risks, since regulation is expected to move in the direction of higher carbon prices, mandatory disclosure of climate-related risks, and border carbon adjustments ().

Innovation will be critical in achieving net-zero emissions by 2050. However, given the path dependent and cumulative nature of knowledge (Nelson and Winter, 1982), an important question is: to what extent is there "intellectual carbon lock-in" whereby society's large history of dirty R&D locks it into further dirty R&D? The question has become all the more important due to the relatively recent technological revolution in hydraulic fracturing that has transformed energy markets.

Models of directed technical change argue that there is hysteresis in the innovation system and that a wholesale pivot to clean innovation would require not only a carbon tax but also directed clean subsidies (Acemoglu et al. 2012; Aghion, et al. 2016). While this broad conclusion is supported by the literature, the calibrations in these models are based on an implicit assumption that clean and dirty technologies do not learn from each other. If clean technology can learn from dirty, then we may be overstating the extent of hysteresis in the system and understating the ability of dirty firms to pivot into clean sectors.

Using the universe of patents granted by the U.S. Patent and Trademark Office (USPTO) from 1976 to 2020, which yields 6 million patents and 55 million connections, we create a measure of "intellectual distance" which assesses the extent to which clean technologies learn from dirty technologies. Clean technologies mitigate greenhouse gas emissions, dirty technologies contribute to emissions and grey technologies are energy-efficient versions of dirty technologies. To identify clean, dirty and grey technologies, we leverage existing classifications schemes in the literature and supplement these with our own tagging efforts (see Supplementary Table 1 and Supplementary Fig. 1). Our measure of intellectual distance is intuitively similar to the concept of "degrees of separation" or the

famous Erdos number, whereby if a clean patent directly cites a prior dirty patent, it has a distance of 1. If a clean patent cites an intermediary patent which in turn cites a dirty patent, the distance is 2. For clean patents that are directly connected to dirty patents, we assess the intensity of the connection by checking what percentage of the clean patent's backward citations refer to dirty patents.

We find that less than one-tenth of clean technologies directly cite dirty prior art indicating that in the vast majority of cases, pivoting from dirty to clean R&D is not straightforward, and there is very likely to be a legitimate case for both carbon pricing and targeted green subsidies to overcome "intellectual carbon lock-in". The average clean patent is connected to a dirty one by three degrees of separation. However, there is a large level of heterogeneity across technologies.

More granular analysis reveals that firms with a history of dirty R&D are well-positioned to diversify into intellectually proximate sectors such as geothermal, carbon capture and storage, and offshore wind, where commonalities such as the need for drilling, pollution control and operating in offshore contexts, creates shared knowhow. These clean sectors are "adjacent" to the hydrocarbon knowledge paradigm. This analysis harkens back to a well-developed literature in export-diversification, which identifies "adjacent sectors" that are feasible diversification options because they build upon the existing knowhow of current export industries (see Hausmann et al. 2014; Hidalgo and Hausmann 2009; Hidalgo et. al. 2007 and, Mealy and Teytelboym 2020 for a green application). However, unlike the economic complexity literature which relies on the co-occurrence of products in an export basket to measure "shared knowhow"/ "adjacency", we leverage patent citations, which are arguably a more robust and direct indicator since they represent causal chains of learning.

Finally, we find that clean technologies like solar photovoltaics (PV), nuclear energy and clean information and communications technology (ICT) are intellectually distant from prior dirty knowledge. While these will not be easy diversification options for dirty firms from a knowledge perspective, they could still be fruitful if one considers other factors that help in diversification such as access to capital, large domestic market, etc.

The direction of clean and dirty innovation

The direction of innovation, even today, is uncertain, as is typified by the reverberating effects of the fracking revolution on the US energy market. Due to path dependencies in the system, it is often easier to advance on an old, mature paradigm than to develop new ones (Tushman and Anderson, 1986; Aghion et al. 2014). Clean patenting grew in the 1970s due to the oil crisis and experienced significant acceleration from the late 1990s and early 2000s (Fig. 1). Rising energy prices drove much of this (e.g. Newell et al. 1999; Popp 2002; Verdolini and Gaelotti, 2011). However, clean patenting peaked right after 2010 and has subsequently fallen due to the burst of the clean tech bubble, challenges in venture capital as a source of finance, and the rise of hydraulic fracturing (Dechezleprêtre, 2017; Gaddy et al., 2017; Popp et al., 2020).

The rise of dirty patenting in 2005, after years of decline was due to innovation in horizontal drilling and fracking (Supplementary Fig. 2b) which resulted in fossil fuel prices falling so much that natural gas became the primary fuel for electricity generation in the US (Popp et al., 2020). This revival of dirty innovation underscores the need to think about "intellectual carbon lock-in" and policy to steer the direction of innovation (Bryan, and Williams 2021).

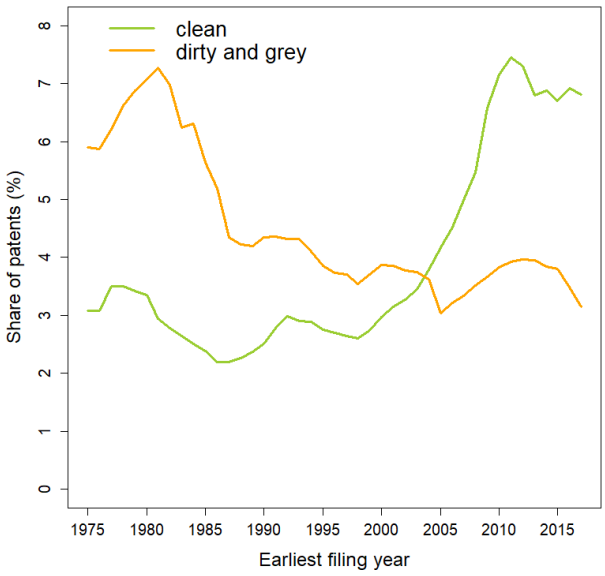


Figure 1. Number of patents as a share of all patents

Notes. The number of clean (green graph) and sum of dirty and grey (orange graph) patents as a share of all US patents by year. Patents are sorted by priority year (i.e. earliest filing year) and counted at DOCDB family-level.

Intellectual distance from clean to dirty technologies

Our distance metric $D_i \in \{1, 2, 3, \ldots\}$ is defined as a clean patent i's minimum citation distance to prior dirty patents within the patent citation network (see Fig. 2 and Methods). When D = 1, the clean patent directly cites a dirty patent. D = 2 indicates that there is an intermediate technology between the clean and dirty patent; D = 3 indicates that there are two intermediary technologies, and so on. The concept can also apply the other way around or for any two classes of technologies. Clean patents that cannot be connected at any distance to prior dirty patents are defined as "unconnected."

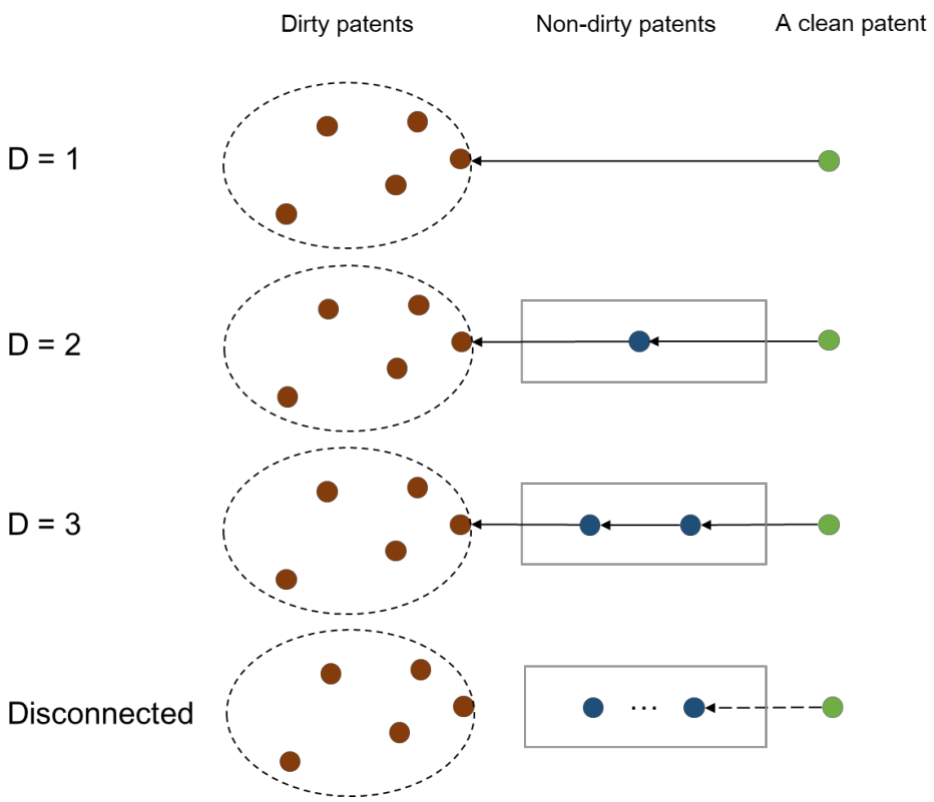


Figure 2. A schematic diagram showing the intellectual distances

Notes. This figure shows how the distance from a clean patent to prior dirty patent is determined. Arrows indicate the citation direction (from citing to cited patents).

As D_i increases, the likelihood that a dirty incumbent can leverage their existing R&D capabilities to diversify into a clean technology becomes smaller (Cohen and Levinthal, 1990; Lane and Lubatkin, 1998). Direct connections from clean to dirty patents comprise relatively a small proportion of the overall connections (7.5%), although the

majority (73%) of clean patents ultimately have some link to dirty prior art (Fig. 3a)[1]. The highest frequency of connection occurs at $D = 3$ (Fig. 3b)[2]. By comparison, the average distance between two randomly selected patents is 8.5 (Mostafavi et al., 2012). The relative proximity between clean and dirty patents is partly attributable to the fact that many have common goals (e.g. generating electricity). Yet the limited proportion of direct connections points towards the fact that it is still not straightforward, in most cases, for entities with a history of dirty R&D to pivot into clean sectors by leveraging the knowledge they already have.

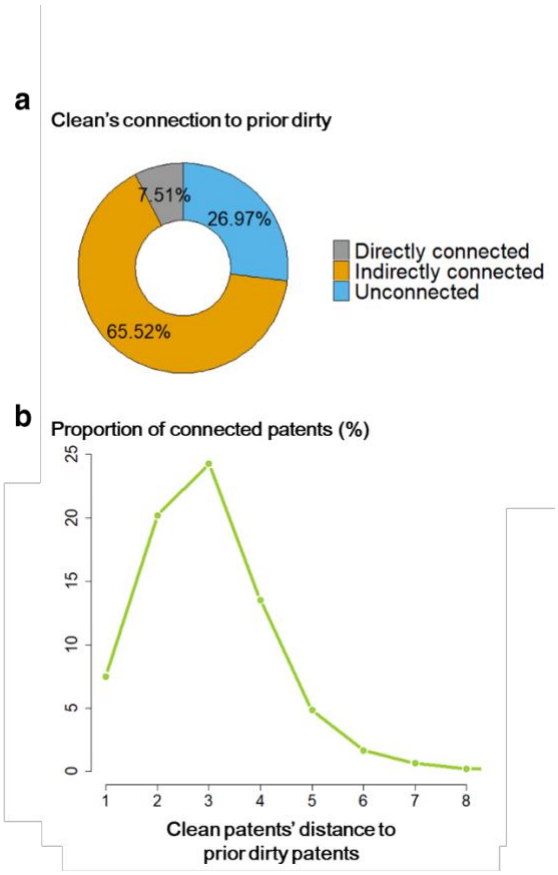


Figure 3. Connectivity from clean to prior dirty patents

Notes. Figures are made based on the directed graph of citation network from clean to previous dirty patents.

[1] Patent citations can be added by both applicant and examiner. Our main analysis does not distinguish between both types of citations because examiner-added citations can be interpreted as "knowledge used by inventors even though the inventors were not fully aware of the existence of relevant prior art" (Berchicci and van de Vrande 2019). As a robustness test, we report results focusing only on the applicant-added citations, which results in a higher proportion of disconnection between clean and dirty technologies (Supplementary Fig. 4).

[2] See Supplementary Fig. 3 for the connectivity from dirty to prior clean patents.

Sectoral patterns of connectivity between clean and dirty technologies

Direct connections

We now investigate *which* clean sectors have a high proportion of patents that directly cite prior dirty patents, as these represent the most straightforward diversification options. As Fig. 4 demonstrates, there is large heterogeneity in the extent to which clean technologies directly learn from dirty ones. Geothermal energy and CCS are amongst the most reliant on the hydrocarbon knowledge paradigm, with at least one-fifth of patents directly citing dirty patents. This makes intuitive sense since geothermal energy relies on geological surveying, drilling techniques, field development, and the construction of wells, pipelines, and other infrastructure, which require knowledge inputs that are commonly used by fossil fuel firms. CCS, for its part, is a complement to coal-fired power plants, gas stations and other point-sources of carbon emissions and naturally has to understand how equipment can be fitted to these.

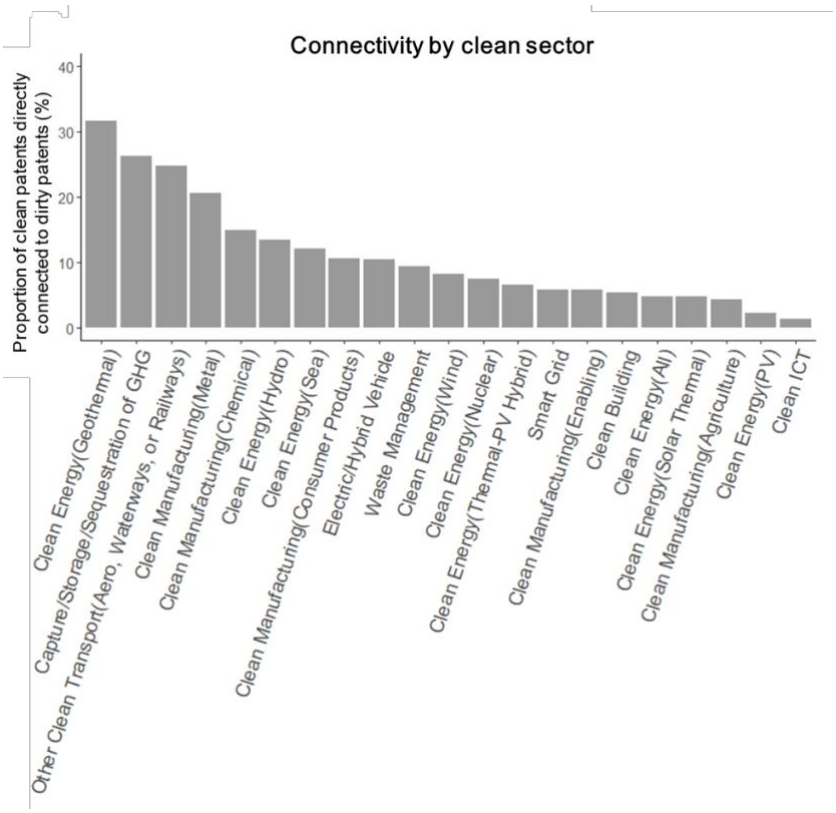


Figure 4. Sectoral details of direct connections

Notes. Each clean sector's direct connections to prior dirty technologies.

These results imply dirty innovators are reasonably well-placed to diversify into clean R&D related to geothermal energy and CCS. However, in practice, diversification requires considering other variables. In the case of geothermal, one has to be close to a viable source, and for CCS one has to consider the "social license" to operate. CCS carries the moral hazard of prolonging carbon-intensive production and as such, has faced backlash from various environmental groups.

Clean innovation in long-distance transport, metals and chemicals is also proximate to dirty knowledge (Fig. 4). Till date, most of it reflects very incremental, process related improvements to energy efficiency to decrease carbon emissions. However, there is also a strand of more radical zero-carbon innovation in these technologies such as zero-carbon steel made from hydrogen (e.g. the HYBRIT project) and "renewable fuels which convert renewable energy sources into chemical molecules for use in various applications" (Dinçer and Bicer 2019). Such innovation is still so nascent that it is unlikely to be sufficiently captured in patent databases. To the extent that incremental innovation crowds out more radical clean innovation, there may be a form of "intellectual carbon-lock in" whereby dirty innovators diversify into adjacent clean sectors, but these are not the ones that drive rapid decarbonisation.

Technologies such as wind turbines, tidal energy, hydro, electric vehicles and waste management are in the middle, with 10-15% of patents being directly connected to dirty technologies. Offshore wind and tidal energy, for example, require knowledge inputs that are common to dirty technologies such as offshore oil. This includes seabed engineering, constructing offshore platforms, placing under-sea cables, under-sea robots and materials that can withstand bio-fouling. This perhaps explains why offshore oil companies like British Petroleum put in aggressive bids for seabed rights in the North Sea to develop offshore wind farms, as they can leverage the capital, skills and knowhow that they have already accumulated (King, 2021). Other elements of wind energy such as onshore wind and components related to smart ICT are relatively more distant from dirty technologies (Popp et al. 2020). It is similar for electric vehicles where some elements such as car design are common to internal combustion engine-based cars, whereas other elements, such as those related to batteries, are different.

The most distant technologies are clean ICT, nuclear and solar PV which all have negligible direct links to dirty technologies and represent a significant shift away from the dirty knowledge paradigm. In these areas, a history of dirty R&D will not confer a comparative advantage. Case study analysis finds that solar energy and other renewable technologies are technologically proximate to ICT innovation (Popp et al 2020). An interesting avenue for future work could involve assessing the intellectual distance between clean technologies and other – more general – sectors of the economy.

However, for the analysis presented above, one may argue that a direct connection between a clean and dirty patent can exist even when only a minor proportion of the clean technology relies on dirty knowledge. To account for this, we assess the "intensity" of the direct connection through the share of dirty backward citations in each clean patent (Fig. 5). Complementing our distance metric with an intensity metric is an important robustness check. Through a systematic assessment, we see that there is a strong positive correlation between the proportion of direct connections and the intensity of the connection (Fig. 5). This indicates that where there is a high proportion of direct citations, as is the case for geothermal and CCS, there is also a strong reliance on dirty technologies. In the case of geothermal, over a quarter of backward citations refer to the hydrocarbon patents.

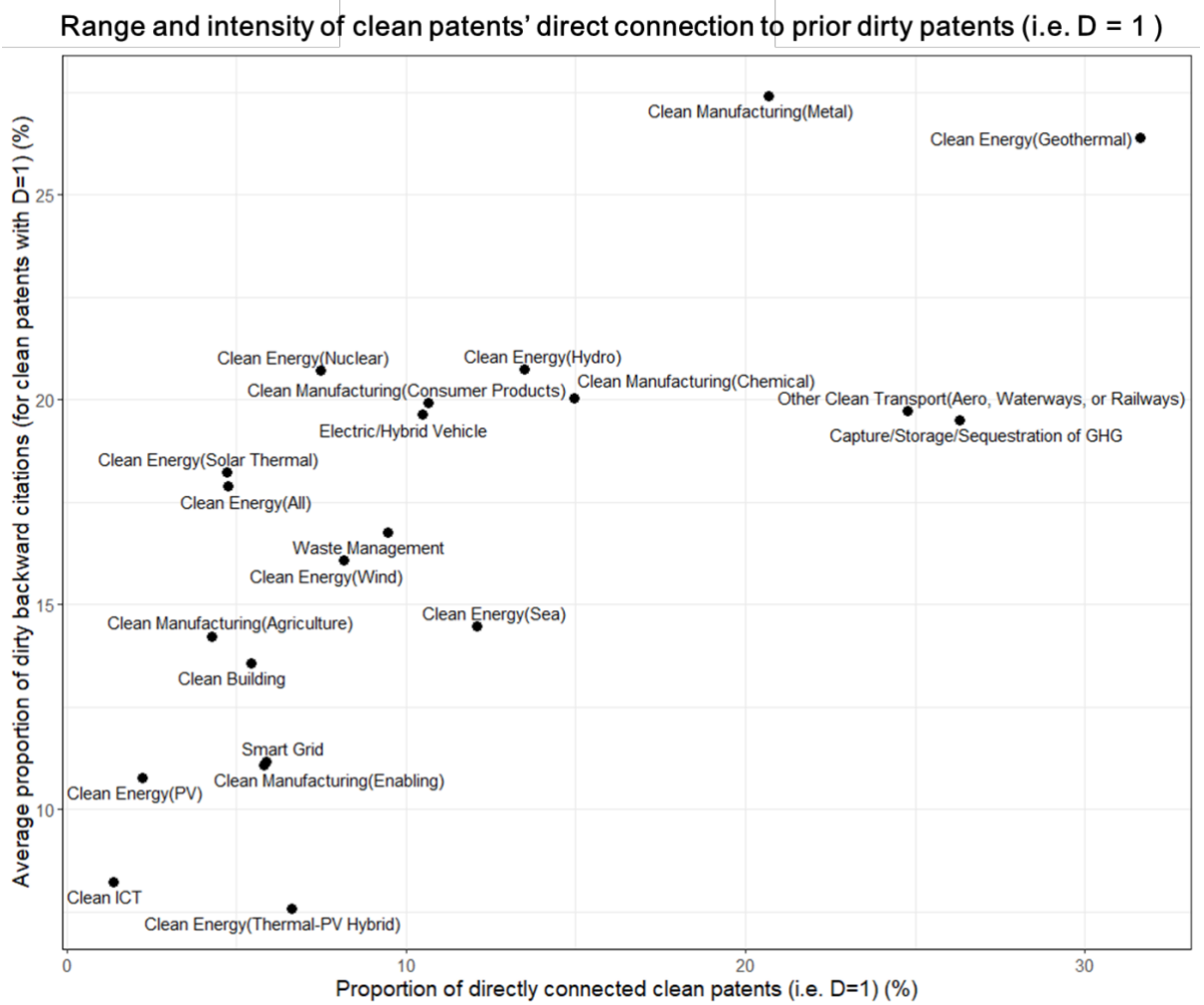


Figure 5. Range and intensity of direct connections

Notes. X-axis indicates the proportion of clean patents that are directly connected to prior dirty patents among the clean patents in each sector (Range of direct connection) Y-axis is the average of the proportion of dirty backward citations among the total backward citations in each clean patent with D = 1 (Intensity of direct connection). Overall, the figure shows that the range and intensity of direct connection tend to be positively correlated.

Indirect connections

Majority of the connections between clean and dirty technologies are indirect (Fig. 3). The distance at which clean patents are connected to dirty patents differs largely by technology, as plotted in Fig. 6. Curves that are shifted to the right depict more intellectually distant technologies and lower peaks indicate a smaller proportion of patents at that particular distance.

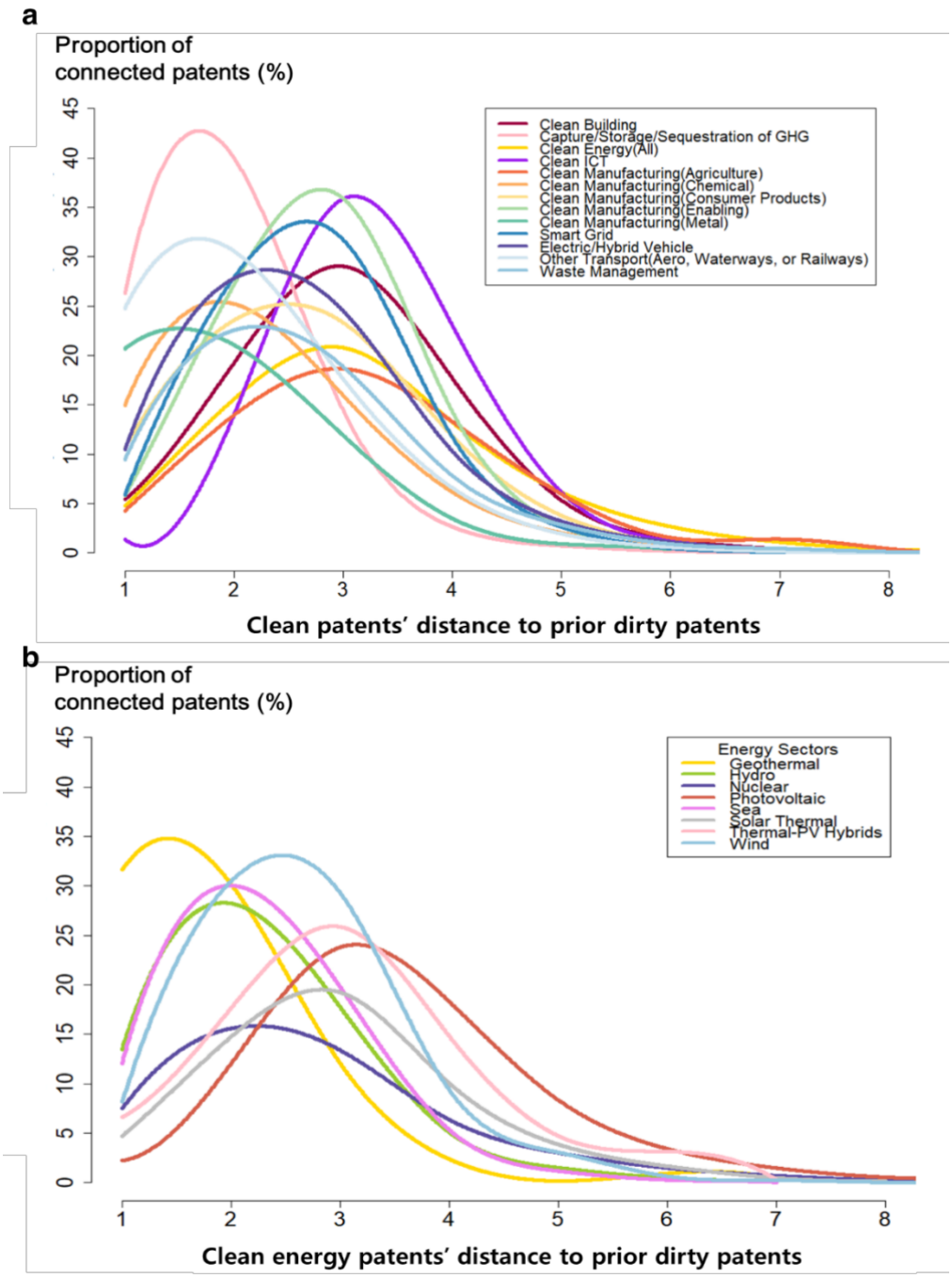


Figure 6. The distribution of clean patents' distance to prior dirty patents

Notes. Sectoral distribution of clean patents' distance to prior dirty patents. Cubic spline interpolation is used to plot the smooth curves. **(a)** For all clean sectors. Clean manufacturing (Enabling) includes manufacturing technologies that can be applied to various domains, with a potential to greenhouse gas emissions mitigation (e.g. greenhouse gas management systems). **(b)** For clean energy sub sectors.

Clean ICT and Smart Grids have a relatively low share of direct connections but a high share of indirect connections at average distance around 2 and 3, indicating that these technologies indirectly build upon the hydrocarbon knowledge paradigm (Fig. 6a). This is not unexpected since there may be IT-related knowledge inputs that are embedded in hydrocarbon technologies that other technologies have built upon which eventually feed into smart grids etc.

In Energy, geothermal is the closet to the dirty knowledge paradigm. The second group is hydro, wind and tidal (sea) energy, which has a high share of indirect learning. The final group is solar and nuclear, which is intellectually distant.

The remaining question is *which* dirty technologies do adjacent clean technologies draw from? Fig. 7 maps clean technologies to connected dirty technologies in a disaggregated manner. Electric vehicles, smart grids, enabling technologies in manufacturing, clean buildings, and clean ICT all learn from dirty transportation but at increasing levels of indirectness (star dots in Fig. 7a). Clean technologies in hard-to-decarbonise sectors, such as chemicals, metals and long-distance transportation share connections with dirty manufacturing and downstream fossil fuel technologies (triangle and diamond dots in Fig. 7a). Geothermal learns significantly from upstream fossil (yellow dots in Fig. 7b).

Grey technologies

Grey technologies either can be a steppingstone towards decarbonisation or can lock-in emissions. For example, in areas where renewable energy is available, building new gas infrastructure may simply exacerbate carbon lock-in even though it is less polluting than coal. On the other hand, if no radically clean substitutes are available immediately, as might be the case in sectors such as iron & steel or cement, grey technologies related to metal recycling or changes in clinker can help incrementally reduce emissions while more radical solutions are being developed (Stern and Valero, 2021). Dirty patents are, on average closer to grey patents than clean patents (compare Supplementary Fig. 3b and Supplementary Fig. 6d). For dirty innovators, the development of grey technologies is therefore a relatively accessible direction of change in contexts where grey technologies can positively mediate the clean transition without exacerbating carbon lock-in.

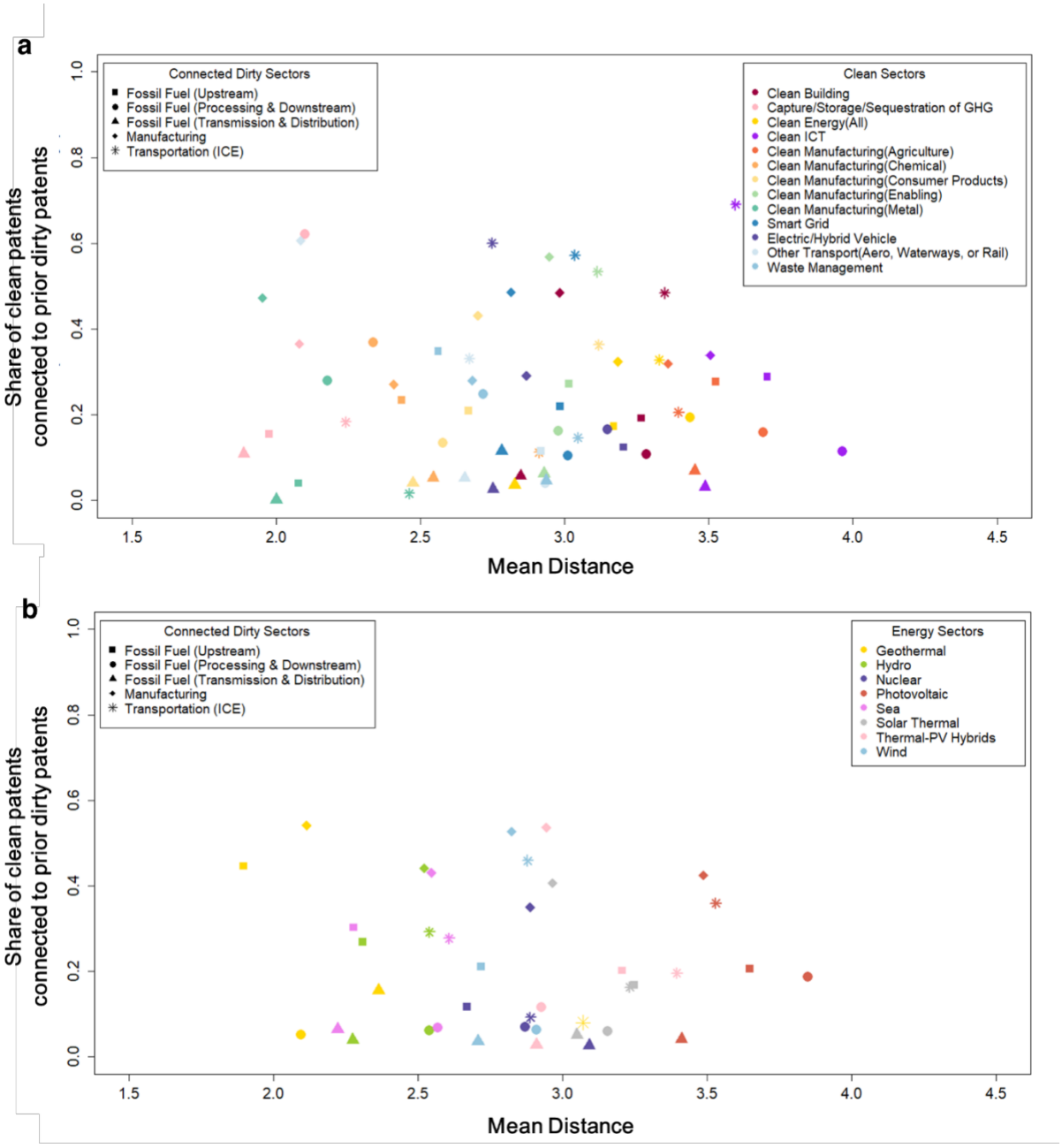


Figure 7. The pattern of sectoral connections from clean patents to prior dirty patents

Notes. X-axis presents the mean of distances from patents in a certain clean sector (indicated by color of dots) to patents in a certain prior dirty sector (indicated by shape of dots). Y-axis presents the share of connected patents in a certain clean sector to patents in a certain prior dirty sector. **(a)** For all clean sectors. **(b)** For clean energy sub sectors only.

Discussion

In this paper we empirically investigate the concept of "intellectual carbon lock-in." The seminal work in directed technical change and the environment argues that due to hysteresis in the innovation system, carbon prices and targeted green innovation subsidies are needed to pivot innovation towards clean technologies (Acemoglu et al. 2012). However, this model implicitly assumes that clean technologies do not learn from their dirty counterparts. We are the first to empirically test this claim, and find that on the whole, clean technologies, particularly those that have been instrumental for the clean energy transition such as solar, are intellectually distant from dirty technologies. Only one-tenth of clean patents directly cite dirty prior art. This lends support for the idea that directed technological subsidies are necessary to pivot the innovation system towards clean technologies.

However, aggregate results conceal areas of high intellectual proximity that offer promising diversification options for hydrocarbon incumbents. In reality, the degree to which clean learns from dirty is nuanced, sector-specific and sometimes indirect. The clean sectors that draw significantly from the hydrocarbon knowledge paradigm include geothermal, carbon capture and storage, and offshore wind. Other areas of clean adjacency include electric vehicles, tidal energy, and more efficient industrial processing. However, some of these learn indirectly rather than directly.

While our results speak to diversification on a knowledge basis, we recognize that there are other factors that contribute to the ease of diversification such as access to capital, access to markets, government support etc. which are not captured in our analysis. Nevertheless, since knowledge is a key input into the production process, understanding the intellectual proximity between clean and dirty technologies is important. Existing literature has focused on whether clean or dirty technologies have higher spillovers (e.g. Noailly and Shestalova 2017; Dechezleprêtre et al. 2014) but this is different from asking how much clean technology learns from dirty and vice versa.

Our analysis also has relevance for countries that are locked into hydrocarbons and are looking for clean adjacent sectors to pivot into to avoid stranded skills, capital and labour. South Africa, India, China and Indonesia, for example, have a long history of coal production and are home to large state-owned enterprises that specialize in mining, drilling, processing

etc. These firms may be well-positioned to transition into adjacent clean sectors by leveraging their current knowhow. This provides a way of limiting the transitional costs of pivoting to clean, since existing skills and capital will find gainful employment in adjacent sectors.

Our approach, in this respect, is similar in spirit of Hausmann et al. (2014)'s complexity analysis, which identifies adjacent sectors for export diversification. However, methodologically, our analysis leverages patent citation data, which is a more robust indicator of shared knowhow than the co-occurrence of products in an export basket. The latter is subject to some degree of inherent randomness whereas patent citations represent a causal chain of learning.

Examining the links between clean and dirty innovation is critical since innovation has cumulative impacts on economic growth and is not directionally neutral. As the fracking revolution revealed, the overall direction of innovation is uncertain and in the absence of policy, it can move in directions that do not support the clean energy transition. The target of net-zero emissions by 2050 necessitates clean innovation, especially in the hard-to-decarbonise sectors. Innovation policy must actively support this aim.

References

Acemoglu, D., Aghion, P., Bursztyn, L. and Hemous, D., 2012. The environment and dire cted technical change. *American Economic Review*, *102*(1), pp.131-66.

Aghion, P., Dechezleprêtre, A., Hemous, D., Martin, R., & Van Reenen, J. 2016. Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, *124*(1), 1-51.

Aghion, P., Dechezleprêtre, A., Hemous, D., Martin, R. and Van Reenen, J., 2016. Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, *124*(1), pp.1-51.

Aghion, P., Hepburn, C., Teytelboym, A., Zenghelis, D., 2014. Path dependence, Innovatio n and the Economics of Climate change, Centre for Climate Change Economics and Policy/G rantham Research Institute on Climate Change and the Environment Policy Paper & Contribu ting Paper to New Climate Economy.

Ahmadpoor, M., & Jones, B. F. 2017. The dual frontier: Patented inventions and prior scientific advance. *Science*, *357*(6351), 583-587.

Berchicci, L., & van de Vrande, V. 2019. Noisy or valuable? The effect of examiner-added citations on firm knowledge flows. In DRUID conference.

Bryan, K. A., & Williams, H. L. 2021. *Innovation: Market Failures and Public Policies* (No. w29173). National Bureau of Economic Research.

Cohen, W. M., & Levinthal, D. A. 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 128-152.

Dechezleprêtre, A., Martin, R., & Mohnen, M. 2014. Knowledge spillovers from clean and dirty technologies.

Dechezleprêtre, A. 2017. "Sustaining Investment in Climate Innovation." Climate Innovation Insights, Series 1.2: Accelerating the Evolution of Climate Innovation Clusters, Climate–KIC. https://www.climate-kic.org/wp content/uploads/2017/03/Insight02_Proof4.pdf

Dechezleprêtre, A., Muckley, C. B., & Neelakantan, P. 2021. Is firm-level clean or dirty innovation valued more?. *The European Journal of Finance*, *27*(1-2), 31-61.

Dinçer, İ. and Bicer, Y., 2019. *Integrated Energy Systems for Multigeneration*. Elsevier.

Gaddy, B. E., Sivaram, V., Jones, T. B., & Wayman, L. 2017. Venture capital and cleantech: The wrong model for energy innovation. *Energy Policy*, *102*, 385-395.

Haščič, I., & Migotto, M. 2015. Measuring Environmental Innovation Using Patent Data 89. pp. 0–1 OECD Environment Working Papers.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., & Simoes, A. 2014. *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. MIT Press.

Hidalgo, C.A. and Hausmann, R., 2009. The building blocks of economic complexity. *Proceedings of the national academy of sciences*, *106*(26), pp.10570-10575.

Hidalgo, C.A., Klinger, B., Barabási, A.L. and Hausmann, R., 2007. The product space conditions the development of nations. *Science*, *317*(5837), pp.482-487.

IEA (International Energy Agency). 2021. Methodology for identifying fossil fuel supply related technologies in patent data.

Jamasmie, C. 2020. *Fossil fuel industry in "terminal decline" — report*. [online] MINING.COM. Available at: https://www.mining.com/fossil-fuel-industry-in-terminal-decline-report/ [Accessed 13 Nov. 2021].

King, I. 2021. *BP blows away bid rivals with big bet on wind farms*. [online] Sky News. Available at: https://news.sky.com/story/bp-blows-away-bid-rivals-with-big-bet-on-wind-farms-12212580 [Accessed 13 Nov. 2021].

Lane, P. J., & Lubatkin, M. 1998. Relative absorptive capacity and interorganizational learning. *Strategic Management Journal*, *19*(5), 461-477.

Mealy, P., Teytelboym, A., 2020. Economic complexity and the green economy. *Research Policy*, 103948.

Mostafavi, S., Goldenberg, A., & Morris, Q. 2012. Labeling nodes using three degrees of propagation. *PloS one*, *7*(12), e51947.

Nelson, R. R., & Winter, S. G. 1982. An evolutionary theory of economic change. Cambridge, MA: Harvard University Press.

Newell, R. G., Jaffe, A. B., & Stavins, R. N. 1999. The induced innovation hypothesis and energy-saving technological change. *The Quarterly Journal of Economics*, *114*(3), 941-975.

Noailly, J., & Shestalova, V. 2017. Knowledge spillovers from renewable energy technolo gies: Lessons from patent citations. *Environmental Innovation and Societal Transitions*, *22*, 1 -14.

Popp, D., Pless, J., Haščič, I. & Johnstone, N. 2020. Innovation and Entrepreneurship in th e Energy Sector (No. w27145). *National Bureau of Economic Research.*

Stern, N., & Valero, A., 2021. Research policy, Chris Freeman special issue innovation, gr owth and the transition to net-zero emissions. *Research Policy*, *50*(9), 104293.

Tushman, M. L., & Anderson, P. 1986. Technological discontinuities and organizational environments. *Administrative Science Quarterly*, 439-465.

Verdolini, E., & Galeotti, M. 2011. At home and abroad: An empirical analysis of innovation and diffusion in energy technologies. *Journal of Environmental Economics and Management*, *61*(2), 119-134.

Author Contributions

Both authors contributed equally to the manuscript.

Acknowledgements

Authors would like to thank Jacquelyn Pless, Cameron Hepburn, Sam Fankhauser, François Lafond, Joris Bücker, Kerstin Hötte, Brian O'Callaghan and INET seminar participants for their feedback. Su Jung Jee acknowledges support from Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A3A03037237). Sugandha Srivastav acknowledges support from the Climate Compatible Growth Programme at Oxford, the Oxford Martin School Programme on the Post Carbon Transition, and the Institute of New Economic Thinking at the University of Oxford.

Methods

Data

We use World Patent Statistical Database (PATSTAT) (Spring 2021 version) to obtain the clean, dirty and grey patents issued by USPTO from 1976 to 2020 and the full patent citation network within the period. To obtain clean, dirty and grey patents as comprehensively as possible, we conduct systematic literature review and combine patent search strategies from complementary sources.

Haščič and Migotto (2015) and Popp et al. (2020) provide patent classification codes for overall clean patents and clean energy patents, respectively. Aghion et al. (2016) suggests the codes for clean, dirty and grey patents in the automobile sector, and Dechezleprêtre et al. (2021) gives the codes for overall clean and dirty patents. To capture a wider array of dirty technologies, we particularly refer to a recent work by International Energy Agency (IEA, 2021), which provides a search strategy based on both cooperative patent classification (CPC) and keywords to find dirty patents associated with fossil fuels' complete supply chain.

To the best of our knowledge, as systematic search strategies for grey patents do not exist yet except for Aghion et al. (2016), which only covers automobiles, we additionally conduct manual sorting of full digits CPCs related to the energy-efficiency improvement of dirty technologies. The sorting of grey CPC codes is conducted within carbon reduction (Y02E), clean manufacturing (Y02P) and clean transportation (Y02T) sectors, which involve sub-codes related to the grey patents (Supplementary Table 1).

The search strategy gives 258,078, 145,753 and 98,224 of clean, dirty and grey patents, respectively (counted at DOCDB family level to avoid redundant counting of same inventions). Patents classified as both clean and dirty ('31,053' patents) are excluded from our baseline analysis to focus on the obvious clean and dirty technologies (see Supplementary Fig. 1). As an indicative illustration: renewable energy and electric vehicle patents are classified as clean, oil and gas patents are dirty, and energy-efficient methods of making steel are grey. The directed citation network consists of 55,188,499 links connecting patents at the family level.

Distance metric

This paper adopts the idea underlying the distance metric used by Ahmadpoor and Jones (2017). The metric was devised to measure the minimum distance between papers and patents across fields to systematically evaluate the degree of applied and basic research nature of each field. Authors defined $D_i \in \{1, 2, 3, \ldots\}$ as a patent *i*'s minimum distance to prior paper within the integrated citation network of papers and patents. If $D_i = 1$ (i.e. patent *i* directly cites one of the prior papers), the corresponding paper and patent are interpreted as representing "patent-paper boundary" where distinction between science and technology is ambiguous.

In our context, D_i is a clean patent *i*'s minimum citation distance to prior dirty patent within the patent citation network. $D_i = 1$ presents clean patents that directly learn from prior dirty patents, while $D_i \geq 2$ indicates clean patents that indirectly learn from prior dirty patents. That is, if $D_i \geq n$ (where *n* is larger than 2), a clean patent *i* must pass through at least *n-1* non-dirty patents within the citation network to reach prior dirty patent(s). Clean patents that cannot be connected to dirty patents at any distance are treated as "unconnected" or "disconnected." The D metric is adapted to conduct additional analyses reported in our manuscript and supplementary material section (e.g., Supplementary Fig. 3 show connectivity from dirty to prior clean patents, Supplementary Fig. 6a and 6b shows connectivity from clean to prior grey patents, and Supplementary Fig. 6c and 6d shows connectivity from dirty to prior grey patents).

Supplementary Tables

Supplementary Table 1. Patent search strategy

	Sector	Sub-sector	CPC	IPC
Clean[1]	Clean Building		Y02B	
	Capture/Storage/Sequestration of GHG (CCS)		Y02C	
	Clean Energy	Geothermal	Y02E 10/10	
		Hydro	Y02E 10/20	
		Nuclear	Y02E 30	
		Photovoltaic	Y02E 10/50	
		Sea	Y02E 10/30	
		Solar thermal	Y02E 10/40	
		Thermal-PV hybrids	Y02E 10/60	
		Wind	Y02E 10/70	
		Others	Y02E 40, Y02E 50, Y02E 60, Y02E 70	
	Clean ICT		Y02D	
	Clean Manufacturing[2]	Agriculture	Y02P 60	
		Chemical	Y02P 20, Y02P 30, Y02P 40	
		Consumer products	Y02P 70, Y02P 80	
		Enabling	Y02P 90	
		Metal	Y02P 10	
	Smart Grid		Y04S	
	Clean Transport[3]	Electric/Hybrid vehicle	Y02T 10, Y02T 90	see Aghion et al. (2016)[4]
		Aero, waterways, or railways	Y02T 30, Y02T 50, Y02T 70	
	Waste Management		Y02W	
Dirty	Energy	Upstream	see IEA (2021)[5]	
		Processing and downstream		
		Transmission and distribution		
	Transport	Internal combustion engine		see Aghion et al. (2016)
	General (Manufacturing)			see Dechezleprêtre et al. (2021)

	Energy[6]	Combustion technologies with mitigation potential	Y02E 20	
Grey[8]	Manufacturing[7]	Technologies related to metal processing/Recycling	Y02P 10/20	
		Technologies related to metal processing/Process efficiency	Y02P 10/25	
		Technologies relating to chemical industry/Process efficiency/Energy recovery, e.g. by cogeneration, H2recovery or pressure recovery turbines	Y02P 20/129	
		Technologies relating to chemical industry/Feedstock/the feedstock being recycled material, e.g. plastics	Y02P 20/143	
		Technologies relating to chemical industry/Reduction of greenhouse gas [GHG] emissions, e.g. CO2	Y02P 20/151/low	
		Technologies relating to chemical industry/Improvements relating to chlorine production	Y02P 20/20/low	
		Technologies relating to chemical industry/Improvements relating to adipic acid or caprolactam production	Y02P 20/30/low	
		Technologies relating to chemical industry/Improvements relating to fluorochloro hydrocarbon, e.g. chlorodifluoromethane [HCFC-22] production	Y02P 20/40/low	
		Technologies relating to chemical industry/Improvements relating to the production of bulk chemicals	Y02P 20/50/low	
		Technologies relating to oil refining and petrochemical industry/Ethylene production	Y02P 30/40	
		Technologies relating to the processing of minerals/Production of cement, e.g. improving or optimising the production methods; Cement grinding/Energy efficiency measures, e.g. improving or optimising the production methods	Y02P 40/121	
		Technologies relating to the processing of minerals/Glass production, e.g. reusing waste heat during processing or shaping/Improving the yield, e-g- reduction of reject rates	Y02P 40/50, Y02P 40/57	
		Technologies relating to the processing of minerals/Production of ceramic materials or ceramic elements, e.g. substitution of clay or shale by alternative raw materials, e.g. ashes	Y02P 40/60	
		Climate change mitigation technologies in the production process for final industrial or consumer products/Manufacturing or production processes characterised by the final manufactured product	Y02P 70/50/low	
		Climate change mitigation technologies for sector-wide applications/Reducing waste in manufacturing processes; Calculations of released waste quantities	Y02P 80/30	
		Climate change mitigation technologies for sector-wide applications/Minimising material used in manufacturing processes	Y02P 80/40	
	Transport	Fuel efficiency of internal combustion engine based vehicles	Y02T 10/10	see Aghion et al. (2016)

Notes. 1. We refer to Haščič and Migotto (2015), Popp et al. (2020) and Aghion et al. (2016) to collect clean patents; 2. Grey manufacturing patents are excluded from the list of clean manufacturing patents; 3. Grey transport patents are excluded from the list of clean transport patents; 4. B06L 11 is transferred to B60L 50/00 - B60L 58/40 (see IPC ver. 2019); 5. Search strategy of IEA (2021) combines both CPC and keyword based search; 6. Authors' interpretation; 7. Authors' interpretation; 8. Grey patents are excluded from the final list of clean and dirty patents. Patents classified as both clean and dirty patents are excluded from our main analysis (see Supplementary Fig. 1).

Supplementary Table 2. Top 10 countries of creating clean, dirty and grey patents

	Clean patents		Dirty patents		Grey patents	
	Country	**Patent counts (Family)**	**Country**	**Patent counts (Family)**	**Country**	**Patent counts (Family)**
1	US	122,515	US	88,319	US	39,660
2	Japan	56,632	Japan	19,377	Japan	28,462
3	Germany	22,622	Germany	13,344	Germany	13,930
4	South Korea	18,755	UK	6,546	South Korea	3,530
5	France	9,230	France	6,465	France	3,422
6	China	8,172	Canada	5,675	UK	3,394
7	Taiwan	7,312	Switzerland	2,177	Canada	2,084
8	UK	6,914	Netherlands	2,081	Italy	1,704
9	Canada	6,472	Sweden	1,669	Switzerland	1,477
10	Switzerland	3,327	Norway	1,529	China	1,244

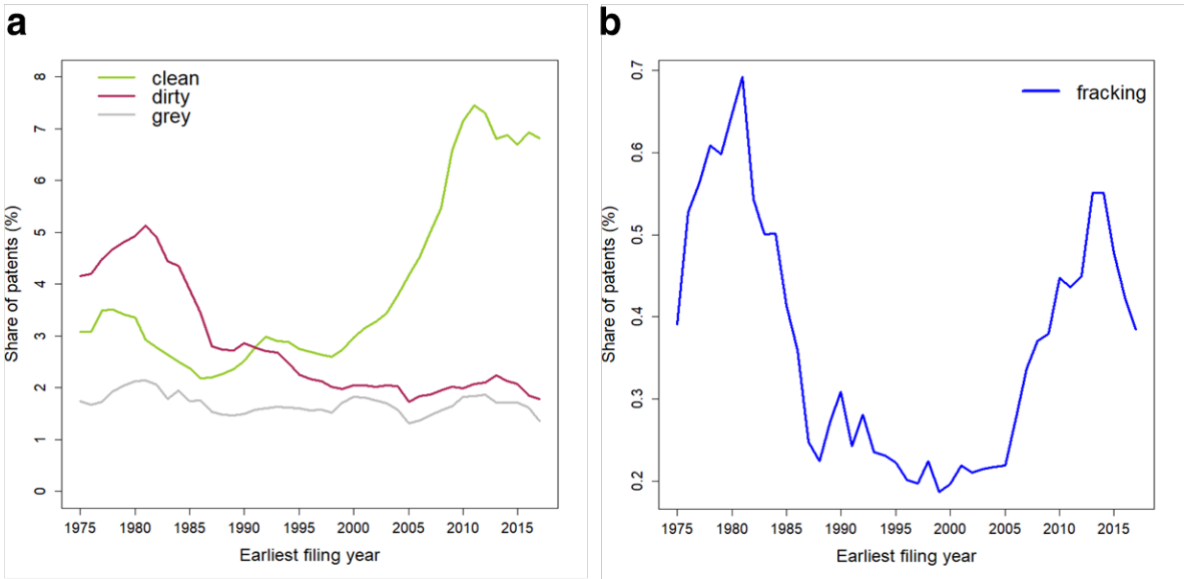
Notes: Patents issued by USPTO during 1976-2020

Supplementary Figures

Supplementary Figure 1. Venn diagram for clean, dirty and grey patents

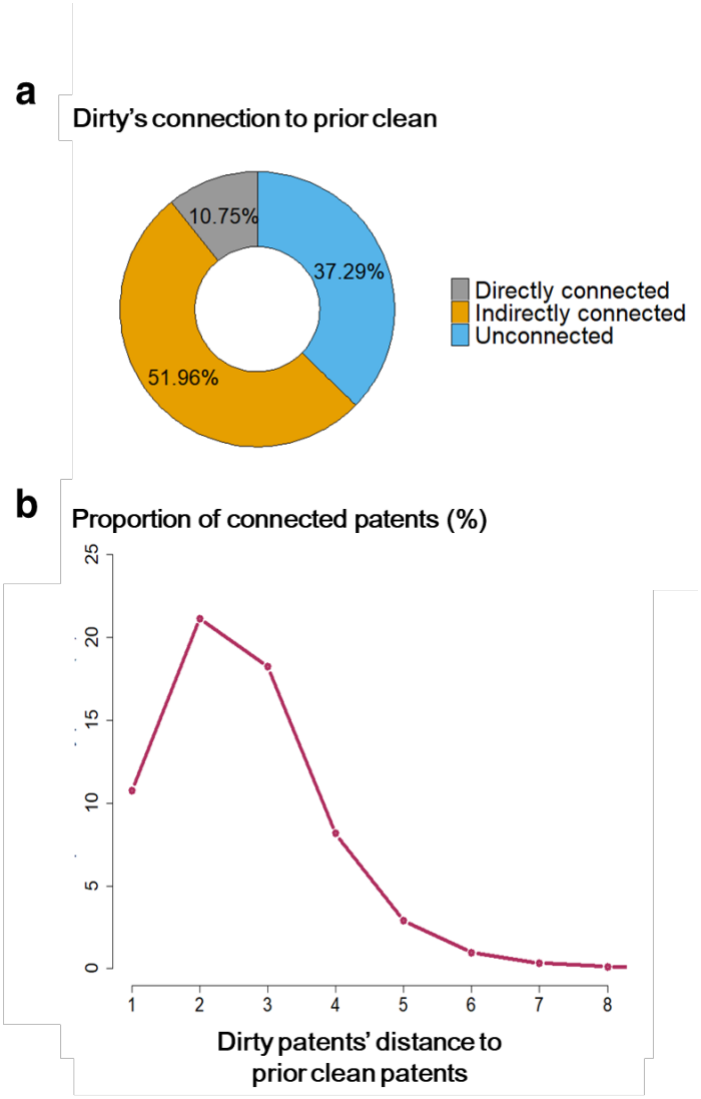
Notes. We regard clean or dirty patents that are also classified grey patents as grey patents. Patents classified as both clean and dirty patents are excluded from our main analysis

Supplementary Figure 2. Number of patents as a share of all patents



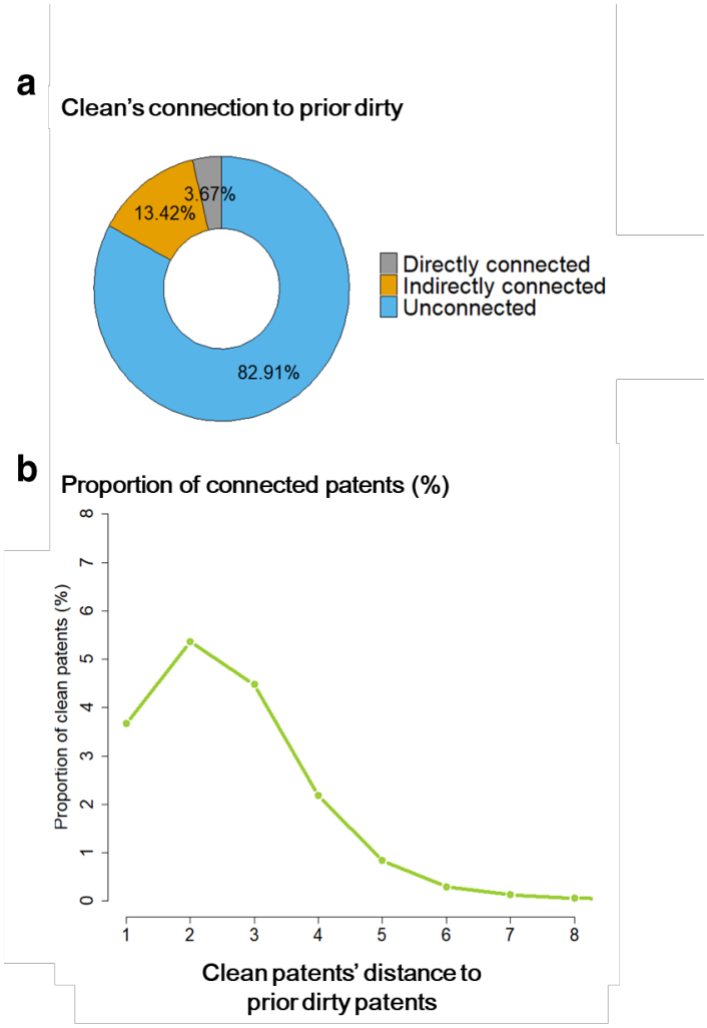
Notes. **(a)** The number of clean, dirty and grey patents as a share of all the US patents by year, **(b)** The number of fracking related patents as a share of all US patents by year. To search fracking related patents, we combine the search strategy for unconventional fossil fuel patents in IEA (2021) and for hydrofracturing patents in Popp et al. (2020). Patents are sorted by priority year (i.e. earliest filing year) and counted at DOCDB family-level.

Supplementary Figure 3. Connectivity from dirty to prior clean patents



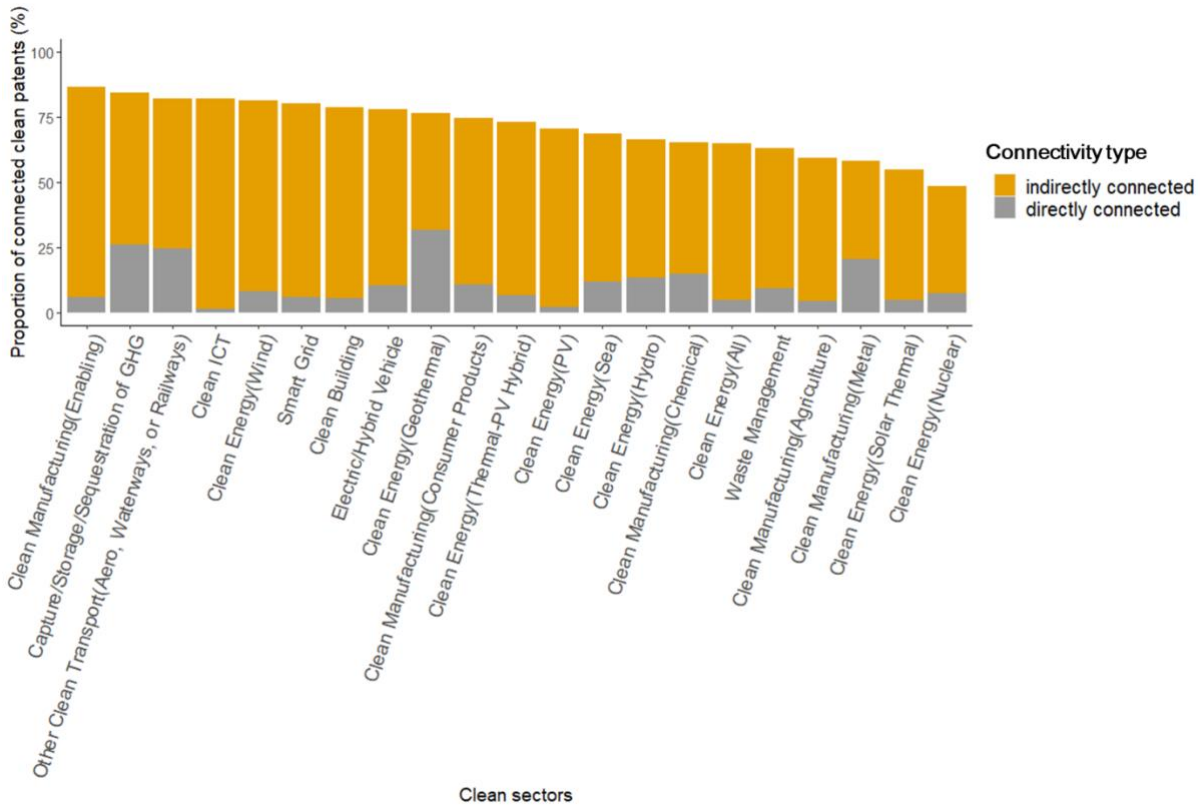
Notes. Figures are made based on the directed graph of citation network from dirty to previous clean patents.

Supplementary Figure 4. Connectivity from clean to prior dirty patents (Applicant citation only)



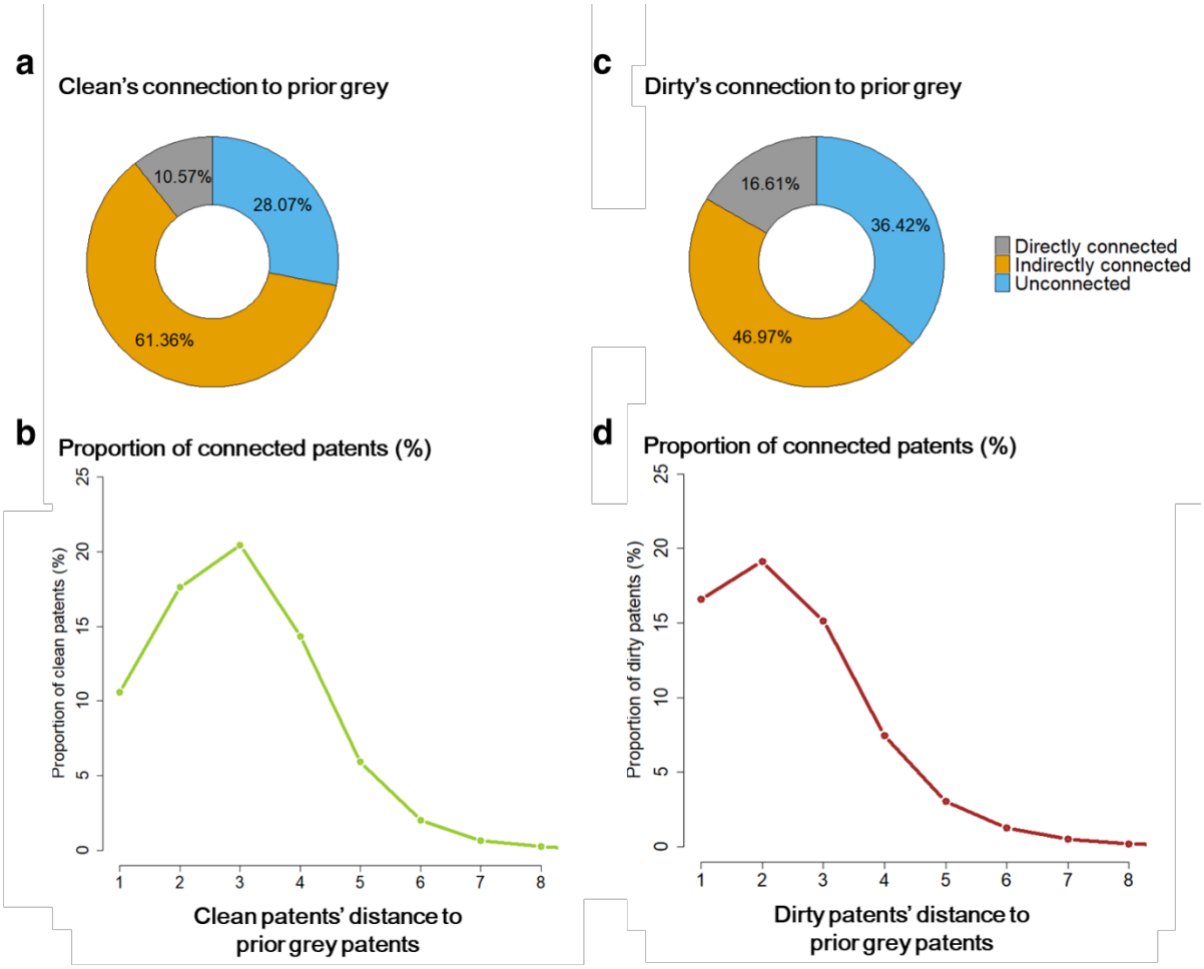
Notes. Figures are made based on the directed graph of citation network from clean to previous dirty patents.

Supplementary Figure 5. Connectivity by clean sector



Notes. Each clean sector's direct (grey) and indirect (orange) connections to prior dirty technologies.

Supplementary Figure 6. Connectivity analysis for grey patents



Notes. Figures **a** and **b** are made based on the directed graph of citation network from clean to previous grey patents and Figures **c** and **d** are made based on the directed graph of citation network from dirty to previous grey patents.