A Penny for your Thoughts*

W. Walker Hanlon† Stephan Heblich‡ Ferdinando Monte§ Martin B. Schmitz¶

May 18, 2022

Please click here for the latest version

Abstract

How do communication costs affect the production of new ideas and inventions? To answer this question, we study the introduction of the Uniform Penny Post in Great Britain in 1840. This reform replaced the previous system of expensive distance-based postage fees with a uniform low rate of one penny for sending letters anywhere in the country. The result was a large spatially-varied reduction in the cost of communicating across locations. We study the impact of this reform on the production of scientific knowledge using citation links constructed from a leading academic journal, the *Philosophical Transactions* and the impact on the development of new technology using patent data. Our results provide quantitative causal estimates showing how a fall in communication costs can increase the rate at which scientific knowledge is exchanged and new ideas and technologies are developed. This evidence lends direct empirical support to an extensive theoretical literature in economic growth and urban economics positing that more ideas can emerge from communication between individuals.

Keywords: Communication, Innovation, Knowledge Flows, Penny Black Stamp
JEL codes: R1, N9, O3

*We thank Bill Collins, Ezra Oberfield, Kevin O'Rourke, Esteban Rossi-Hansberg, Nico Voigtländer, and Ariell Zimran for useful discussions and comments. We also thank seminar participants at Georgetown, Marquette, the Max Planck Institute for Innovation and Competition, NYU Abu Dhabi, USC, and Vanderbilt as well as the 2019 BSE Summer Forum Workshop on Geography, Trade, and Growth and the 2019 UEA meeting. Finally, we thank Riley Alexander, Devin Bissky Dziadyk, Griffin Carroll, Chloe Chiles, Carly Genovere, Henry Goldberg, Rowan Isaaks, John Landry, Jessica Moses, Rachel Norsby, Nicholas Richmond, Andrew Warren, Calee White, and Niel Yocom for excellent research assistance. The usual disclaimer applies.

†Northwestern University, Department of Economics, NBER and CEPR

‡University of Toronto, Munk School of Global Affairs and Public Policy and Department of Economics and NBER

§Georgetown University, McDonough School of Business, NBER and CEPR

¶Vanderbilt University, Department of Economics

1 Introduction

The exchange of knowledge and new ideas is central to many facets of economic activity. For example, the sustained increase in standards of living enjoyed by modern economies is often explained by the creation and diffusion of ideas (Romer, 1986; Lucas, 1988, 2009; Buera & Oberfield, 2020). Further, the costly nature of ideas exchange generates agglomeration forces that contribute to explain the existence of cities (Marshall, 1890; Duranton & Puga, 2004; Davis & Dingel, 2019).

The fundamental premise in these views is that new ideas are more likely to emerge when communication between individuals is less expensive. Although such a premise is natural, isolating the causal impact of communication costs on the circulation and creation of knowledge has proven challenging. This consideration is particularly true in modern environments, where individuals can exchange knowledge through a wide variety of close substitute channels, and many confounding factors limit the scope of the analysis.

In this paper, we empirically investigate the effect of communication costs on the creation of new ideas. To circumvent some of the challenges found in modern environments, we turn to Great Britain from 1830 to 1849, a historical setting where letters, the primary form of long-distance communication, had few close substitute channels. In 1840, the introduction of the "Uniform Penny Post" (Postage Act, 1839) spurred a country-wide communication revolution. This reform, spearheaded by Rowland Hill, replaced an expensive system of distance-based postage with a single, low, uniform charge of one penny for a standard-sized letter. Because the post was the primary form of long-distance communication at this time, this reform had a major impact on communication costs. Also, since the new uniform charge replaced a distance-based postage system, this reduction was spatially varied.

We hypothesize that the change in communication costs may have affected the production of both basic science and applied technology. A long line of economics research, dating back at least to Nelson (1959), Schmookler (1966), Mansfield (1980), and Griliches (1986), highlights the contributions of both basic science and applied research to economic growth. More recently, this distinction has been formalized in work by Aghion & Howitt (1996) and Akcigit *et al.* (2021). This literature suggests that both types of knowledge development contribute to long run growth, though it also recognizes that it may take many years, or even decades, for new basic science insights to be incorporated into applied research leading to productivity improvements. Since both basic and applied work involve the development of new knowledge, both may be influenced by the extent to which ideas can be shared and discussed. Hence, our empirical analysis aims to study the impact of reduced communication costs on both basic and applied research.

To measure the impact of the reform on scientific progress, we focus on articles and citations in the leading British scientific journal of the period, the *Philosophical Transactions of the Royal Society of London*. The primary challenge in constructing this new outcome measure is identifying and geolocating both the article authors and those they cite. Geolocating these individuals requires manually searching for and reviewing biographical sources on each individual, an extremely labor-intensive process. Through this process we were able to identify and geolocate more than 90 percent of article authors and those that they cite for each year from 1830-1849, providing a decade of observations on either side of the reform.[1]

To study how citations between scientists were affected by the reduction in communication costs, we need to measure the change in postage costs for letters flowing between any pair of scientists. In order to do so, we digitized and geolocated a list of 618 post offices and over 1,600 sub-post offices operating in England, Wales or Scotland just before the reform, and the postal road network through which each post office connected to the others. Each scientist is then matched to the nearest post office, either directly or through a sub-post office, allowing us to calculate the change in postage costs between each pair of post office locations induced by the reform.

A key feature of this "citation dataset" is that the citations reflect bilateral flows of knowledge inputs. This allows us to adopt standard analysis methods from the trade literature which estimate the impact of a reduction in trade costs (or in our case communication costs) on flows while controlling for origin-time, destination-time, and location-pair fixed effects. Using this method, we find evidence of a substantial increase in citations between scientist pairs that experienced a greater reduction in communication costs as a result of the postal reform. Specifically, our results indicate that the introduction of the Uniform Penny Post eliminated around 70% of the decay of citations associated with distance-dependant communication costs across locations. This provides a first piece of evidence on the impact of communication costs on scientific knowledge flows.

To measure the impact of communication costs on applied research and technology development, we examine the impact of the reduction in communication costs on the development of new patented technologies. This outcome reflects a second dimension through which communication costs may have affected the development of new ideas, and one that has a more direct connection to productivity growth. To conduct this analysis, we geolocated all addresses reported in the thousands of patents filed between 1830 and 1849. There is one important difference between our patent data analysis approach and what we do with the citation data. Unlike the citation dataset, our patent dataset reflects location-level outcomes, rather than bilateral flows, so it does not admit the analysis approach that is possible with the citation data.[2] Because patents are a location-level outcome, we need to construct a location-level measure of treatment due to the reform. To do so, we follow the "market access" approach of Donaldson & Hornbeck (2016), but with two important differences. First, the cost of communication is determined by postage rates. Second, these rates are based on the length of the postal routes which mail carriers followed at the time. Naturally, this "letter market access" measure is related to a location's market access, a measure or proximity to other economic centers that accounts for transportation costs. Hence, our preferred patent data analysis specification includes a standard market access measure, calculated over the joint waterways, rail and road network, as a control.

To implement this research design, we need to estimate the increase in towns' letter market access

[1] This share is out of those cited scientists that were alive at the time that the article citing them was published.

[2] Patent citations are not available in the setting that we study.

caused by the reform, for which elasticities of letter flows to travel costs, postage costs, population and market access are needed. We recover these elasticities using data on letters sent to London from hundreds of post towns during one week in 1838. This preliminary step also provides direct evidence that postal costs had a large impact on the volume of letter flows.

Using a simple difference-in-difference strategy, our patent data analysis shows that locations that experienced a greater increase in letter market access produced more patents after the postage reform, controlling for each location's market access as well as other potentially important confounders, such as the location's own population and county-by-decade trends. Our results are robust to a variety of estimation approaches. Our patent data analysis provides a second complementary and consistent piece of evidence on the impact of the fall in communication costs on the exchange of scientific knowledge and the production of new ideas.

By providing direct evidence on the influence of communication costs on science and innovation, our results contribute direct empirical support to the link between communication among individuals and the diffusion and creation of new ideas. This link is at the core of a large literature in economic growth (Romer, 1986; Lucas, 1988, 2009; Buera & Oberfield, 2020) and urban economics (Marshall, 1890; Duranton & Puga, 2004; Davis & Dingel, 2019).

Our results fill in a missing piece between several existing strands of work. On the one hand, there are a number of studies that focus on isolating the impact of communication costs or communication disruptions on other economic outcomes, including Jensen (2007), Goyal (2010), Aker (2010), Allen (2014), Koudijs (2014), and Steinwender (2018). Our study differs from this work in that we focus on the impact of changing communication costs on scientific knowledge and technology development, two outcomes that are of particular importance for economic growth.

Our study is also related to existing work looking at how changes in trade costs influence innovation rates. Agrawal *et al.* (2017), for example, estimates the impact of highways on innovation, while Catalini *et al.* (2020) studies the impact of a fall in airfares. An important distinction between work in this area and our study is that changes in highways or air transport can affect both the cost of transporting goods (or people) as well as the cost of communication. A novel feature of our study is our ability to isolate the impact of changes in communication costs from the effect of broader changes in transport costs.

Another related strand of research uses patent data or academic citations to infer the existence of knowledge flows related to science or invention (Jaffe *et al.*, 1993; Thompson, 2006; Murata *et al.*, 2014). Existing evidence suggests that communication costs likely play an important role in inventive activity.[3] However, it is difficult to establish a direct causal relationship using these methods because studies in this area typically do not observe plausibly exogenous changes in communication costs, and because increased communication via one channel may in part reflect reduced communication via the many alternative, close substitute channels. This makes it hard to isolate the role of communication from other omitted

[3]Additional evidence shows that inventive activity tends to be geographically agglomerated, and more so than manufacturing activities in the same industry (Audretsch & Feldman, 1996; Carlino *et al.*, 2012).

local factors, as well as other impacts of proximity such as reduced transport costs.

A particularly productive strand of work focuses on academic research, where some of the identification issues faced by studies of the broader economy can be overcome. For example, Waldinger (2011) uses the expulsion of scientists by the Nazis to provide evidence on localized peer effects. Another paper that is even closer to our study is Agrawal & Goldfarb (2008). In that paper the authors examine the impact of a very specific reduction in communication costs between universities – the adoption of Bitnet, a precursor of the Internet – on inter-university collaboration in engineering. Their results indicate that the reduction in communication costs increased collaboration between university researchers.[4] There are two important differences between our study and Agrawal & Goldfarb (2008). First, our citation data analysis takes advantage of information on bilateral knowledge flows (citations) and bilateral variation in the change in communication costs, which allows an analysis approach that addresses many potential identification concerns. Second, our evidence on the impact of reduced communication costs on scientific knowledge is complemented by our evidence on the development of new technologies, which are likely to be particularly important for improving economic growth.

Our study also contributes to existing work looking at the impact of postal systems on economic development. One closely related study, Acemoglu *et al.* (2016), uses the presence of post offices as an indicator of state capacity and then shows that this measure is correlated with subsequent patenting activity. While these results suggest a link between the presence of post offices and innovative output, they do not attempt to isolate the importance of communication costs from other aspects of state capacity. Another closely related paper, Rogowski *et al.* (2019), uses a combination of cross-national and U.S. county-level data on the extension of the postal system and finds that greater access to the postal system was associated with faster development (as indicated by national GDP or county level farm values, manufacturing output or capital investment). Relative to Acemoglu *et al.* (2016), our study provide more direct causal estimates of the impact of reduced communication costs, through the postal system, on innovation. Relative to Rogowski *et al.* (2019), we offer both a more cleanly identified analysis approach as well as evidence on how communication costs affected scientific and technology development, rather than braoder economic development.[5]

Our paper is also closely related to a contemporaneous working paper by Abhay & Xu (2022). Their study looks at how a bureaucratic reform to post offices in the U.S. in the late nineteenth century affected mail flows and patenting patterns. There are four important differences between our study and theirs. First, our focus is on the impact of communication costs on knowledge production, while they are interested in the impacts of improved bureaucratic efficiency.[6] Both types of reforms are worthy of

[4] Other recent studies in this area include Belenzon & Schankerman (2013) and Boudreau *et al.* (2017).

[5] Another working paper, by Feigenbaum & Rotemberg (2014), also looks at the impact of postal access, using the expansion of postal services in the United States through rural free delivery. They find that this expansion impacted production patterns, but they do not study the impact on innovation.

[6] The experiment considered in Abhay & Xu (2022) does not affect the distance-based cost of communication, and so it cannot speak directly to the issue that we focus on.

study, and so we view these as complementary results. Given our focus, we borrow tools developed for studying the impact of trade costs in order to provide an analysis that speaks directly to the importance of communication costs, as a function of distance, for knowledge production. Second, our use of changing bilateral communication costs, together with the bilateral citation data, allows us to take a particularly strong identification approach, also borrowed from recent studies in international trade, in which we control for origin-period, destination-period, and dyad fixed effects. This approach deals with many potential identification concerns. Third, we examine the impact of a change in communication costs on both basic science and applied technology development, both of which benefit from the exchange of knowledge (as we show) and both of which contribute to economic growth. Fourth, our study focuses on a setting with few alternative methods of long distance communication apart from posting letters. Later in the nineteenth century, alternative methods such as the telegraph became more commonly used. Thus, we view our setting as providing a particularly clean environment for examining the impact of communication costs.

Finally, our paper improves our understanding of a key event in British economic history. The introduction of the penny post provides a particularly interesting example of how an institutional reform can contribute to sustaining technological progress and economic growth: economic historians such as Joel Mokyr have argued that knowledge exchange played a critical role in facilitating technological development during the Industrial Revolution (Mokyr, 2005a). That the reform of this particular institution mattered should not be surprising, given that during the nineteenth century the post office was almost certainly the branch of national government that individuals were most likely to encounter in their everyday lives.[7]

The rest of the paper is organized as follows. Section 2 provides background information on the postal reform. We present our data in Section 3. Section 4 establishes the relationship between postage costs and the volume of letter flows. Our main analysis is in Section 5, followed by a concluding discussion.

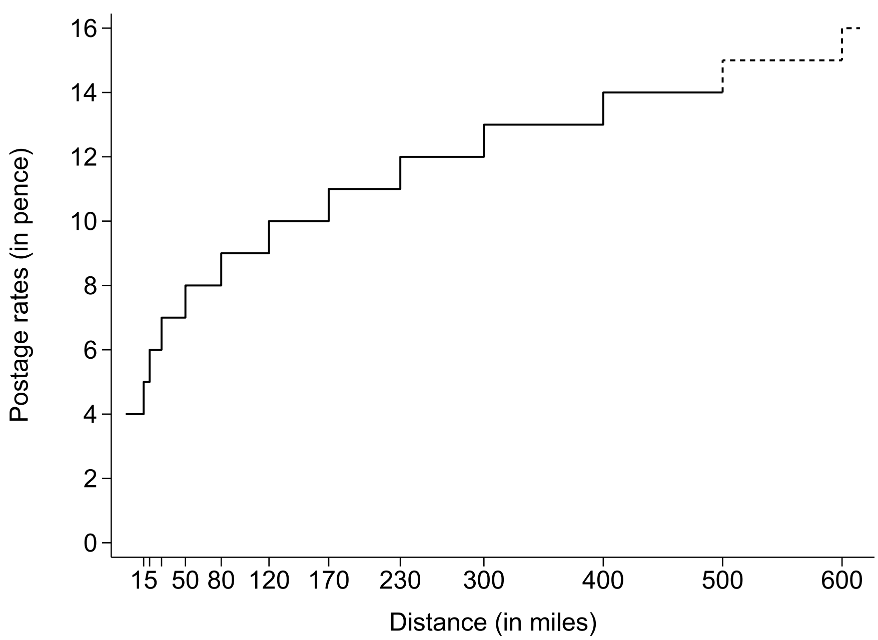
2 Background

In the early nineteenth century, posting letters was the primary means of long-distance communication. Scientists, engineers, and other inventors were often heavy users of the postal system. The surviving correspondence of Michael Faraday, for example, a prominent English scientists working on electromagnetism, comprises over 4,900 letters (James, 1991-2011). So important and voluminous were these correspondences that in many cases they provide the primary record that we have of the lives of scientists and inventors during this period. Letters exchanged between scientists and inventors were often packed with scientific knowledge, technical information, and questions. As the optical scientist and photographic inventor David Brewster wrote to William Henry Fox Talbot in 1837, "My last letter was so crammed with Science, that I could not find a corner to ask your aid in a question of Literature..." (Schaaf, 2021).

[7] For an eloquent exposition of this point, see Acemoglu *et al.* (2016).

However, in the 1830s there was a great deal of dissatisfaction with the postage system in Britain. This was due in part to the fact that, prior to 1840, one of the primary aims of the Post Office was to raise revenue for the government through the use of its monopoly power. Reflecting this aim, postal rates had been repeatedly raised in the early nineteenth century in response to the revenue needs created by the Napoleonic Wars.[8] Not only were costs high, but the system of distance-based rates was complex. Postage was based on the carrier's journey, the cost was dependent on weight and the number of sheets, and the postage was typically paid by the recipient. The postage varied discretely at specific distance thresholds, with wider bands for longer distances. Figure 1 shows the cost function as reconstructed from original documents. This all added to the expense of sending a letter. High expenses also meant lower letter flows, which resulted in less frequent deliveries.

Figure 1: Postage cost of a one-page letter, pre-reform



Note: This figure shows the postage cost for a one-page letter before the postal reform. The dashed line indicates that for every additional 100 miles an additional penny was being charged.

In the late 1830s, Rowland Hill became a leading advocate of post office reform. In 1837, he published

[8] MPs and other high-level government officials were exempt from paying for postage because they had what were called "franking privileges" that allowed them to send ten and receive 15 letters per day at no cost (Postage Act, 1795). These individuals were also known to post letters for some friends. The Postage Act of 1839 ended these franking privileges and introduced uniform one-penny postage for everyone.

his famous pamphlet, *Post Office Reform, its Importance and Practicability,* in which he argued for the introduction of a flat postage rate independent of distance for all letters up to 14 grams, and the introduction of prepaid postage indicated by a stamp on the letter.[9] Hill argued that reducing postage rates would actually increase revenues, because the quantity of letters sent would offset the lower price per letter, while at the same time simplifying the system could also allow the Post Office to handle more letters at a lower cost.

The pamphlet was less popular among Post Office officials who feared increasing costs, but a House of Commons Select Committee led by Robert Wallace, MP was favorable. The Committee reports 320 petitions containing 38,709 signatures in support of Hill's plan (Wallace, 1837-38). As a result, the Penny Postage Bill was passed in July 1839 and a uniform penny postage was officially launched on 10 January 1840.[10] From this day on, "... a Letter not exceeding half an ounce in weight [could] be sent from any part of the United Kingdom, to any other part, for One Penny ..." (General Post Office, 1840, p.1).[11] Only five months later, the introduction of the *Penny Black* stamp, the first adhesive postage stamp worldwide, concluded Britain's transition to the first modern postal system.

The reform dramatically reduced the cost of long-distance postage, leading to a rapid increase in the volume of letters sent. The cost of sending a letter of three sheets from London to Edinburgh, for example, dropped from 39 1/2 pence to 2 pence. This was a large decrease, even for the relatively well-off scientists and engineers in our sample: the decrease in the cost of a three-sheet letter sent from London to Edinburgh was equal to a decrease from around 10-20% of a professor's average daily wage to just 0.5-1%.[12]

From 1839 to 1840, the number of letters posted in Great Britain more than doubled, from 73 million to 151 million, and the volume reached 312 million by 1850 (General Post Office, 1856, p.56). This dramatic increase can be seen in Figure 2, where we present estimates of letter flows that we have constructed.[13] Since few alternatives existed for long-distance communication, such an increase in the volume of letters represents a net increase in communication flows within the country. Greater letter volume also facilitated

[9]Much earlier, in 1680, William Dockwra and his partner Robert Murray established the London Penny Post which was, however, restricted to postings within London. Other local penny posts existed within other parts of the country and exclusively served the local communities within their small coverage areas. The reform we study was focused on inter-city postage, a feature reflected in our analysis.

[10]Some MPs were initially skeptical that Hill's plan would generate sufficient revenue and called for a higher uniform rate. The resulting Uniform Fourpenny Post launched on 5 December 1839, led to a volume increase of 28 percent within less than a month, convinced skeptics of the uniform penny rate, and quickly became the victim of its own success (see Hill, 1840, Coase, 1939).

[11]In contrast to the pre-reform period, this weight limit meant that the lowest rate also applied to two-sheet letters.

[12]The calculation of the lower bound relies on Colquhoun (1806). For the upper bound, a salary of £300 per year is used based on a figure listed in the Oxford Dictionary of National Biography for a chaired professor at the University of Cambridge (see Chapman, 2011).

[13]In England and Wales, the number of mailed letters per capita increased from about 4 in 1839 to about 17 in 1849. These estimates are obtained by dividing the letter flows for England and Wales in 1839 and 1849 by the 1841 and 1851 census population data.

an increase in the frequency of deliveries, which improved the convenience of using the post (see Select Committee on Postage, 1843, pp.258-261).[14] Thus, the reform triggered a communication revolution that allowed people across Britain to exchange ideas and access knowledge at low costs.

Figure 2: Letter volumes over time

Note: This figure shows the number of letters circulated in Great Britain from 1832 to 1849. Details on the construction of this figure can be found in Appendix B.1.

Contemporary sources indicate the benefits that those working in the scientific and technical spheres enjoyed from their improved ability to communicate cheaply across distant locations. The following quotes from a report by the Select Committee on Postage (1843) illustrate this. Evidence taken from Mr. F. Baring discusses the social advantages and states that he is "in possession of various letters, showing some [...] great advantages to literature, science, and friendly union, evinced by the transmission of scientific specimens, evinced, too, by the production of works and the formation of even large societies,

[14] Appendix Figure A1 shows the gross revenue, costs of management and net revenue for the study period 1830–1849. As to be expected, we see s sharp decline in revenue in the year of the reform, 1840. Subsequently, the increasing volume of letters increases gross revenues but since costs increase as well, net revenue remains below the pre-reform level in the decade after the reform.

to the existence of which, as their authors and promoters assure me, the establishment of the penny rate was an essential condition" (p.12). He substantiates this with a letter that Professor John Henslow, a friend and mentor of Charles Darwin, wrote to Rowland Hill in 1843:

> *Dear Sir,—The observation to which you refer in one of my letters to the farmers of Suffolk, respecting the advantages of the penny postage, relates to a scheme of experimental co-operation for securing the rapid progress of agricultural science [...] The practicability of such a scheme depends entirely upon the advantages offered by the penny postage [...] I have circulated 100 copies of a printed schedule, and could have circulated more if I had had them, containing directions how the proposed experiment should be tried. The mere suggestion of this scheme has involved me in a correspondence which I never could have sustained if it had not been for the penny postage. To the importance of the penny postage to those who cultivate science I can bear most unequivocal testimony [...].*

In a similar vein, Mr. E. Tennent mentions his communication with Mr. Manby, the secretary of the Society of Civil Engineers, in the report who states, "that the collection and diffusion of scientific information has been vastly extended by the penny postage." (p.13) Together, these quotes nicely summarize why we would expect this sudden decrease in communication costs to have a marked effect on science and technological progress.[15]

In addition to the postal reform that we study, there were other major changes in transportation that our study will have to deal with. By far the most important of these was the expansion of the railway, following the introduction of the first passenger railway between Liverpool and Manchester in 1830. We will be careful to control for the expansion of the railway system in our analysis. The telegraph was also introduced, starting in 1844, but it was initially just used as a signaling system for the railways so it is less of a concern for our study.[16]

3 Data

3.1 Measuring treatment

Because the reform we study lowered postage between any two points in the country to a low uniform rate, to measure the change in the postage rate due to the reform we simply need to estimate the rate in the pre-reform period. Because the pre-reform rate was distance-based, this requires that we construct the network of postal offices and post roads. Using original sources we have traced out the postal road

[15] The preserved individual correspondence of several scientists and the fact that one third of the male and half of the female population in Britain was still illiterate in 1840 suggest that scientists benefited disproportionately from the reform.

[16] The public initially made little use of the telegraph lines because of their limited availability and high cost (Fava-Verde, 2018). Only in the mid-1850s did the public begin using telegrams for private and business communication.

network (which is not the same as the turnpike network, since not all roads were post roads) and connected them to a newly digitized and geocoded list of 618 post offices.[17] Appendix C contains the details of this step. Figure 3 presents our digitized map of postal roads and post offices.

Figure 3: Post roads and post towns in Great Britain

(a) Post Towns and Sub-Post Towns

(b) Post Towns and Post Roads

Note: The maps show England and Wales. Panel (a) shows post towns (black dots) and sub-post towns (grey crosses) observed around 1838. Panel (b) shows post towns (black dots) and postal roads digitized from maps around 1838.

Essentially all of the large towns in England, Wales and Scotland had a post office, and most of the scientists and patentees that we study were located in one of these post towns. However, some were located in smaller towns, villages, or rural areas. To link these outlying patentees and scientists to our set of post towns, we have digitized and geolocated a more detailed list of over 1,600 sub-post offices, each of which is linked to a post office through which letters mailed at the sub-post office would have flowed. For scientists or patentees outside of the post towns, we link them to their nearest post or sub-post town and then to the main post town that their sub-post town was associated with.

[17] See Select Committee on Postage, 1838, pp.153-165.

The post towns will be our primary unit of analysis. Using the postal route network, we calculate the distance between every pair of post towns. We then use this distance together with the distance-based pricing scheme used in the pre-reform period documented above to calculate the bilateral cost of sending letters between locations before the reform. Since the reform lowered this cost to a low uniform rate, we can use this pre-reform cost to compute the size of the cost reduction induced by the reform.

An important feature of the postage costs in the pre-reform period is that they were a step function of distance. So, for example, to send a single-sheet letter fifteen miles in the pre-reform period cost 4 pence, but for distances from 15 to 20 miles the cost was 5 pence, for 20 to 30 miles it was 6 pence, and so on.[18] At larger distances the bands were wider, so for distances from 80 to 120 miles the cost was 9 pence but the 10 pence band stretched from 120 to 170 miles. This structure, together with the fact that letters traveled along a specific set of postal routes, implies that the postal costs used to construct our key treatment variables are less correlated with bilateral distance. This binned cost structure is a useful feature for our analysis, particularly when we look at the patent data.

3.2 Measuring scientific knowledge flows

We measure scientific knowledge flows using articles published in the *Philosophical Transactions of the Royal Society of London*. First published in 1665, the *Philosophical Transactions* was the premier British scientific journal during our study period. A general-interest journal, it published articles across all branches of science. Articles appearing in the journal had to have been presented at a meeting of the Royal Society (Fyfe et al., 2017). They also went through a review process that involved (typically) two Society Fellows providing referee reports to the editor, generally the Society Secretary, with the authors given an opportunity to revise accepted papers in response to comments (Fyfe et al., 2017).

Our analysis requires citation, location, and biographical data for all scientists who published or were cited in the *Philosophical Transactions* in the study period.[19] The citing practices in the first half of the 19th century necessarily differed from today's. Instead of citing specific publications, authors cited individual scientists and described the cited scientists' particular body of work that they used and built on in their articles, whether or not such work appeared in the *Philosophical Transactions*. The main reason for that is simple. The authors could not expect that readers had access to all publications. The citations in the *Philosophical Transactions* take the form of capitalized last names. Often, but not always, titles such as Prof., Dr., or Mr. are included. Authors typically introduced individuals when they could expect that readers would not be familiar with them.[20] An example of a citation from a 1840 article in the *Philosophical Transactions* is reported in Figure A2 in the Appendix.

[18]Letters with more pages faced higher rates but a identical step function increase in the cost at the same distance bands, as shown in Appendix Table D1.

[19]Appendix E provides additional details on the data collection process and addresses co-authorship.

[20]These short introductions often described the relationship between the author and the cited individual or stated the cited individual's institution or location.

To construct our data set, we began by collecting all of the 443 articles published in the journal from 1830-1849. We then disambiguate the authors of all articles using biographical information from the *Oxford Dictionary of National Biography*, the *Fellow Directory of the Royal Society*, and other similar sources. Finally, we use these sources as well as additional sources such as the 1841 census and city directories to find and record the modal geographic locations of all authors in the publication year(s) of their articles.

After identifying and geolocating all authors of the 443 articles, we limit our attention to the 389 articles that were published by authors located in England, Wales, or Scotland.[21] We then follow a systematic approach, discussed in detail in Appendix E.1, to identify and geolocate all scientists and inventors who are cited in this set of 389 articles. We record and geocode the modal locations of these cited individuals in the publication years of the corresponding articles.[22]

Identifying the cited scientists solely based on their last names, titles, and the article content, in particular around the respective citation, provided for an extremely challenging and resource-intensive data collection.[23] The geolocating process was even more labor-intensive than the identification and disambiguation process as it required an even more detailed review of available biographical information and the search and documentation of additional sources.[24]

After uniquely identifying individuals, we find that our 389 articles include 2,611 citations to scientists who were living at the time the article was published. Out of this group, we were able to geolocate the cited scientists for 2,587 citations, 1,251 of which referred to a scientist living in England, Wales or Scotland, with more than 90 percent of citations geolocated in every year. Table E3 provides annual and period-specific summary statistics for all citations that we extracted from the 389 articles. The fact that we were able to identify and geolocate such a large fraction of scientists is due to the fact that most of them were reasonably prominent individuals for whom at least some surviving biographical information was available, as well as the very labor-intensive manual approach that we applied.

To provide a sense of the type of studies covered by our *Philosophical Transactions* data, we have manually categorized each article into broad categories of scientific fields. Based on article titles, each article was classified into one, and up to three, topics. Table 1 describes the distribution of articles across scientific fields. Clearly, our articles span a broad set of different scientific activities, ranging from basic

[21] Appendix E.2 provides additional information on both sets of articles. Summary statistics are available in Tables E1 and E2.

[22] The modal location in the publication year of the article may differ from the modal location at the time when the cited, sometimes unpublished work was produced, but going back in time to link scientists' particular bodies of work to their previous publications is infeasible. In any case, any error in this location measure due to this difference is unlikely to be correlated with treatment and so is unlikely to affect our estimates.

[23] We constructed an algorithm for identifying the cited scientists, instructed our research assistants how to use the algorithm, ensured that they followed the algorithm, frequently provided feedback to the research assistants, and conducted regular quality checks. Appendix E contains details about this step.

[24] We provided our group of research assistants with detailed instructions on the location data entry and carefully checked the resulting data for accuracy.

Table 1: Topics in Philosophical Transactions, 1830-50

Topic	Share of articles	Share of pages	Topic	Share of articles	Share of pages
Anatomy/Medicine	0.154	0.145	Geology	0.041	0.065
Meteorology	0.113	0.085	Mechanics/Physics	0.027	0.027
Magnetism	0.104	0.110	Botony	0.026	0.021
Electricity	0.092	0.099	Unclassifiable	0.017	0.032
Chemistry	0.092	0.080	Navigation	0.011	0.007
Zoology	0.078	0.081	Timekeeping	0.011	0.018
Geography	0.075	0.063	Archeology	0.010	0.009
Astronomy	0.059	0.077	Metallurgy	0.008	0.005
Light & Heat	0.055	0.057	Sound	0.006	0.005
Enginering/Applied science	0.046	0.041	Demography	0.005	0.002
Mathematics	0.043	0.038	Units of measurement	0.004	0.004
Scientific Equipment	0.043	0.041			

Note: This table shows the distribution of articles published in the Philosophical Transactions across broad categories of scientific fields in our sample.

science to more applied work. One take-away from this table is that many of the scientific topics covered in our citations data are focused on basic science topics, such as mathematics, or other topics, such as zoology, geology, or meteorology, that were unlikely to lead to patented inventions in the short run. Of course, this does not mean that such basic science was not valuable.

Direct correspondence through the post is one way that one scientist may have learned about the work of another. Other channels include reading about their work in scientific publications like the *Philosophical Transactions*, or correspondence through a third party. However, simply learning about existing work was only one step in the research process that led to a new scientific discovery. A review of the correspondence undertaken by scientists during this period reveals that new discoveries were often preceded by extensive communication between scientists. So if cheaper postage facilitated the flow of communication, we would expect to see more articles with citations across longer distances in the post-reform period for two reasons: first, because it was easier for individuals to learn about the work of other scientists who were further away; second, because more frequent correspondence aided in the development of new ideas leading to a publication in the *Philosophical Transactions*.[25]

[25] It is worth noting that, in addition to analyzing bilateral citation data, it is in principle also possible for us to conduct a location-based analysis of the appearance of new scientific articles using the same methods that we will apply to the patent data. While this is possible in theory in our setting, in practice the relatively small number of articles published during our study period means that it is not feasible to obtain sufficient power for this type of analysis. Thus, we focus our attention on the richer information available in the bilateral citation data.

3.3 Measuring the impact on technology development

A natural complement to our analysis of the sharing of more basic scientific knowledge is an analysis of the development of useful technologies. Both scientific knowledge and technological development are likely to make an important contribution to economic growth, though over different time horizons. These activities are also clearly related to one another, though there is an ongoing debate over how much basic science contributed to technological development during the Industrial Revolution.[26]

In order to study how the reduction in communication costs affected the development of new technology, we take advantage of patent data. Patent data provide a rich source of information for understanding technological development during our study period (and continuing on to today). Though patent data are not without their limitations (see, e.g., MacLeod *et al.* (2003) and Moser (2012)), their many advantages have led them to be widely used in studies seeking to understand patterns of technology development. In particular, patents provide a relatively rich set of information covering a large number of new technologies and they are subject to a set of clear and consistent (and in our study period, fairly stable) incentives.

One notable feature of our setting is that filing a patent was a very expensive process. Sullivan (1989) calculates that patent fees in 1830 were at least four times the value of the average output per person in England. This very high cost likely reduced the number of low-quality patents filed during our study period, though it may mean that some useful technologies were never patented. As a consequence, we expect that our patent data include a select and generally higher-value subset of technologies developed during our study period.

For consistency with our citation analysis, our patent analysis focuses on the period from 1830-1849. We focus on England and Wales, which shared a common patent system.[27] Patent laws were stable during this period, though a major change in patent law did take place just after the end of our study period, in 1852.[28] It is useful to note that patents may have been increasing in attractiveness in the 1830s due to a set of court decisions that shifted the case law in favor of protecting patent holders (Bottomley, 2014). To the extent that this improved the attractiveness of filing a patent for inventors across all locations, this will be dealt with as part of our difference in difference analysis strategy.

The patent data that we use were compiled by the British Patent Office (Woodcroft, 1854) as part of the 'Titles of Patents of Invention'. These data were digitized by Nuvolari & Tartari (2011), and we have geocoded them using a combination of the Google geocoding API and manual location searches.[29] We then link each patent to the nearest post town, either directly or through a closer sub-post town. Details

[26] Contributions to this debate include Landes (1969), Rosenberg (1974), Mokyr (2002), Khan (2018), Jacob (2014), and Kelly & Ó Gráda (2020).

[27] Other parts of the U.K., such as Scotland and Ireland, had separate systems requiring separate applications, but the number of patents in these systems was much smaller than the number filed in England and Wales.

[28] For a comprehensive discussion, we refer the interested reader to a recent book by Bottomley (2014) as well as earlier work by Dutton (1984) and MacLeod (1988).

[29] We are in debt to Alessandro Nuvolari for sharing the digitized list of patents with us. For a more detailed discussion of Woodcroft's work, we refer the interested reader to Nuvolari & Tartari (2011).

of this allocation procedure are provided in Appendix F. The result is a list of 7,172 patents filed between 1830 and 1849, each assigned to a post town (our unit of analysis) in England or Wales.[30] Figure A3 in the Appendix illustrates how the number of patents evolves over our study period.

3.4 Travel Costs

We are interested in the effect of a decrease in communication costs on the creation of new ideas. One potential concern is that the emergence of the railway network during this time might bias the estimated effect of decreasing communication costs because it facilitated in-person communication (or other trade flows). To disentangle the effect of in-person communication from postal communication, we calculate the minimum travel costs between all locations on a combined network of (i) coastal waterways; (ii) inland waterways and canals observed in 1830; (iii) turnpike roads observed in 1830; and (iv) newly added railway lines and stations for each year between 1830–1849.[31] Figure 4 shows two example networks, one at the beginning of our time window in 1830 before the construction of the railway network and one at the end of our time window in 1849 when the backbone of the railway network had emerged.

To calculate travel costs, we connect the different transport networks at access points. Railway stations provide access to railway lines; ports provide access to coastal routes; and turnpike roads, navigable rivers and canals can be accessed anywhere. We further assume that travelling on water is 3 times more costly than travelling by rail and travelling on land is 4.5 times more costly than travelling by rail. These weights correspond to the weights estimated in Donaldson (2018) with 19th-century data. Using these cost parameters, we can calculate the minimum travel costs between all locations. This will allow us to control for the steady decline of in-person communication costs and focus on the abrupt change in postal communication costs, i.e. letter market access. The following section discusses our measure of letter market access in detail.

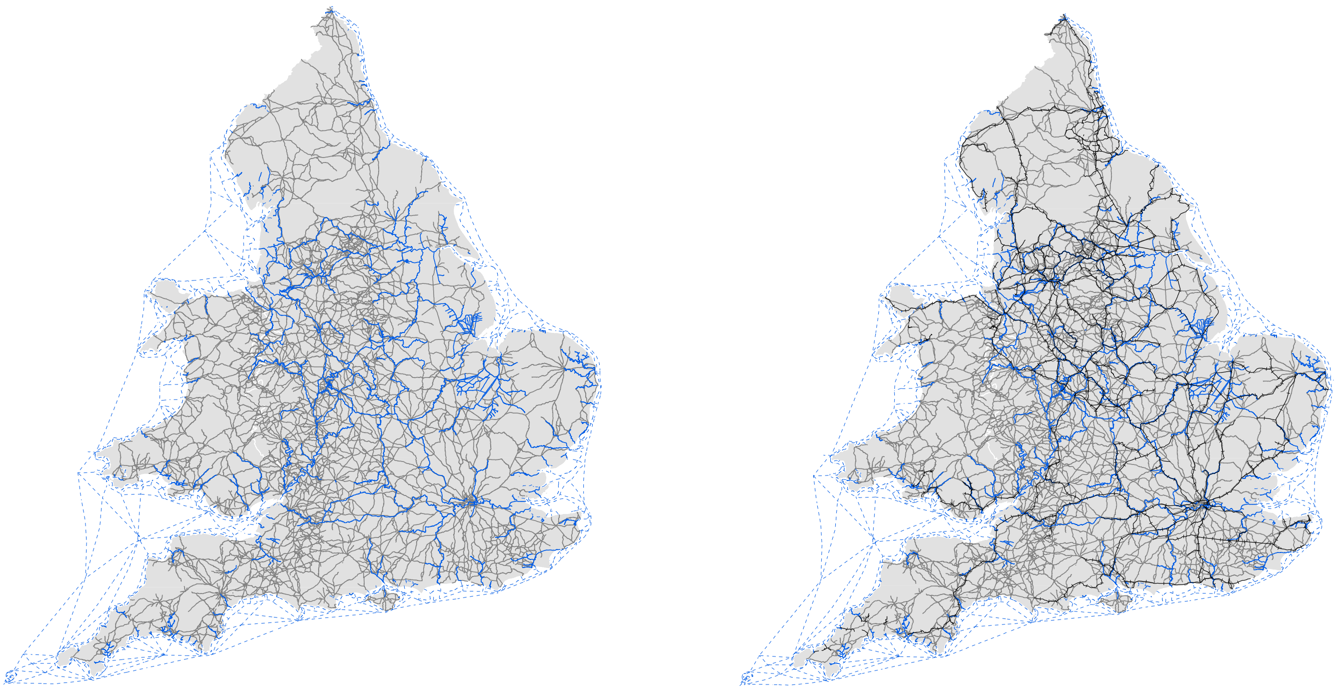
3.5 Letter flows

As part of our analysis, we want to study the impact of the treatment on bilateral letter volumes. Unfortunately, no matrix of bilateral letter flows exists for most of our study period. However, we were able to locate letter flows from a set of 579 post towns to London for one week in 1838 from the *Report of the Select Committee on Postage* of 1838. Since these data come from prior to the reform, they can be used to study how variation in both postage cost and distance to London affected the volume of bilateral

[30] For some patents in our database, the patent agent is listed in place of the inventor. Many of these were patents for inventions that were developed outside of the U.K. Since patent agents were located in London and we do not consider London in the analysis, this is not relevant here. Moreover, we drop all communicated patents, i.e. patents that were communicated from abroad and thus not subject to the knowledge flows we are interested in.

[31] Rosevear *et al.* (2017) created a GIS of turnpike roads, Satchell *et al.* (2017) created a GIS of inland waterways, and Alvarez-Palau *et al.* (2019) created a GIS of ports. Railway lines were derived from a railway atlas by Cobb (2006). Martí-Henneberg *et al.* (2017) created the GIS of England, Wales and Scotland railway stations 1807-1994. See also Bogart *et al.* (2022).

Figure 4: Travel Network in 1830 and 1849



(a) Travel Network in 1830 (b) Travel Network in 1849

Note: The maps show England and Wales separated by white border lines. Panel (a) shows coastal waterways (blue dashed line); 1830 inland waterways (blue line); and 1830 turnpike roads (gray lines). Panel (b) adds all railway lines built until 1849 to the map (black lines).

letter flows. In addition to the volume of regular letter flows, these data also include flows of "privileged" letters, those sent by individuals, such as MPs, that had what were called franking privileges that allowed them to send letters at no cost, as well as flows of newspapers, which were subject to a different cost system. While our analysis will focus on regular letters, flows of newspapers and privileged letters will be useful as control variables.

4 Postage costs and letter flows

Before analyzing the relationship between communication costs and knowledge flows, it is useful to establish the relationship between letter flows and postage costs. This analysis serves two purposes. First, it provides a key piece of evidence that communication costs were indeed strongly limiting communication flows. Second, our measure of a location's treatment in the patents analysis relies on estimates of the relation between letter flows, postage and travel costs, and location-level characteristics that will be recovered in this exercise.

Let L_n be the total number of letters originating from location n. We posit that residents of location n choose to send letters to any location i following a gravity specification of the form:

$$L_{ni} = \frac{A_i d_{ni}^{-\gamma} c_{ni}^{-\eta}}{\sum_{i' \neq n} A_{i'} d_{ni'}^{-\gamma} c_{ni'}^{-\eta}} L_n \tag{1}$$

In this expression, A_i is a parameter that captures the general propensity of location i to receive letters from anywhere; d_{ni} is the travel cost between locations n and i, which captures the fact that bilateral individual relationships are less likely to exist—and hence, bilateral letter flows are lower—when travel between two places is more expensive; c_{ni} is the monetary cost of sending letters from n to i; and the denominator captures the set of outside communication opportunities available to residents in n, which we can think of as a measure of *letter market access* (LMA):

$$LMA_n = \sum_{i' \neq n} A_{i'} d_{ni'}^{-\gamma} c_{ni'}^{-\eta}. \tag{2}$$

The letter market access of an origin location grows with $A_{i'}$ and falls with travel and communication costs when γ, $\eta > 0$: hence, a larger LMA indicates that location n's residents have a better opportunity to communicate 1) at low postage costs 2) with closer places 3) which tend to be more attractive communication destinations.

Note that communication within town ($i' = n$) is excluded from our measure of letter market access, which is a slight departure from the trade-inspired gravity equation for letters. We make this choice for three related reasons. First, personal interactions were a closer substitute for letters sent *within* a town but not over longer distances *between* towns. Second, historical evidence suggests that within-town communication was often used to convey short messages to *coordinate* personal interactions, rather than convey information *as a substitute for* personal interaction. Third, the type of information conveyed in letters is explicit, *codified* knowledge, whereas coordination for personal interactions facilitates the exchange of explicit as well as *tacit* knowledge. These three considerations imply that the gravity model we use for letter flows to London might not apply to distances that can be also covered by short individual travel: both the link between letter flows and our three explanatory variables, and the link between information content and letter flows are of a different nature within versus between towns. Hence, in our analysis, we focus on access to long-distance communication and exclude communication within towns from LMA.

Taking logs of eq. 1, we obtain

$$\ln L_{ni} = -\gamma \ln d_{ni} - \eta \ln c_{ni} + \ln A_i + \ln L_n - \ln \sum_{i' \neq n} A_{i'} d_{ni'}^{-\gamma} c_{ni'}^{-\eta} \tag{3}$$

In standard gravity regressions, the terms $\ln A_i$ and $\ln L_n - \ln \sum_{i' \neq n} A_{i'} d_{ni'}^{-\gamma} c_{ni'}^{-\eta}$ would be absorbed by destination and origin fixed effects, respectively. As we discussed in Subsection 3.5, bilateral letter flows were never collected for this time period, and we only have the number of regular letters sent to London from 579 origins for one week in January 1838.[32] This constraint has two implications.

First, it requires us that we fix $i = London$, so that the letter-destination fixed effect A_i will enter the regression as a constant. Second, while we have measures of d_{ni} and c_{ni} between location n and London, we do not observe the total volume of letters sent by a post town to all destinations, L_n, nor the attractiveness parameters $A_{i'}$. Since we cannot control for these terms with letter-origin fixed effects (we would have as many fixed effects as observations), we need to specify a relation between letters, attractiveness, and location-level observable characteristics. We assume that the propensities of a place to send or receive letters are given by $L_n = \beta_0 P_n^{\beta} MA_n^{\kappa}$ and $A_i = \beta_1 L_i = \beta_1 \beta_0 P_i^{\beta} MA_i^{\kappa}$. In these expressions, P_n is the population of location n[33] and MA_n is the *market access* of location n, a measure of its access to other economic centers as proxied by the inverse travel costs–weighted proximity to all population centers in Great Britain. More formally, market access is defined as

$$MA_n = \left(\sum_{i \neq n} P_i d_{ni}^{-\theta} \right) \tag{4}$$

where P_i is the population of location i, d_{ni} is the travel cost computed over the combined waterways, rail, and turnpike networks in 1838 between locations n and i in 1838, as described in Section 3.4, and θ is the "trade elasticity", which captures the effect of space-related frictions on bilateral economic activity. Note that we exclude the own population of location n since such population will always be present as a separate regressor in our analysis. Our assumptions about L_n and A_i are natural: a location is more likely to send or receive communication when it has a large population and when it has better access to other large economic centers. Eq. 2 then indirectly implies that location n has greater *communication* opportunities (higher LMA) when it is closer to places that have larger *economic* opportunities, as measured by their population or their access to other large markets.

With these assumptions, we can turn to describe the process of recovering our key elasticities. We adopt a two-part procedure. In the first part, we estimate the impact of travel costs, population, and market access (γ, β, and κ) using only variation within postage cost bins, by including fixed effects for each of the discrete postage rate bins. In the second part, we estimate the impact of postage costs on letter flows, η, by comparing locations within narrow windows on either side of the points where postage costs jump. This procedure allows us to leverage the sharp discontinuity implied by the step-wise nature

[32] Summary statistics for all the variables used in our analysis of letter flows are available in the Appendix, Table B3.

[33] The locations in this analysis are post towns. Population is the registration district population for the district in which the post town is located, obtained from census data. Districts are fairly small and can be thought of as encompassing the local labor market associate with the post town. In a few cases, one registration district has more than one post town. In those cases, we assume that each post town serves an equal fraction of the population of the district.

of the cost of sending letters pre-reform.

In particular, in the first part we use all our 579 observations to estimate

$$\ln L_{ni} = \alpha_0 - \gamma \ln d_{ni} + \beta \ln P_n + \kappa \ln MA_n - \phi \ln \sum_{i' \neq n} P_{i'}^{\beta} MA_{i'}^{\kappa} d_{ni'}^{-\gamma} c_{ni'}^{-\eta} + \alpha' X_n + \varepsilon_n \tag{5}$$

This expression follows from eq. 3 after substituting our assumptions for L_n and $A_{i'}$. Here, α_0 is a constant, d_{ni} is the travel cost between post towns n and i computed over the combined waterways, rail, and turnpike networks in 1838, and c_{ni} is the cost of exchanging letters pre-reform. The terms P_n and MA_n are the 1838 population and market access of the registration district in which the post town is located. The term ε_n captures the sum of classical measurement error in the regressors. The X_n term includes a set of ten region fixed effects and postage cost brackets fixed effects. The postage cost brackets fixed effects more flexibly control for the regressor $\eta \ln c_{ni}$ in eq. 3. Since the cost for sending letters is constant within brackets, any remaining relationship between travel costs and letter flows is independent of the cost of sending letters.

In the second part, we directly exploit the fact that postage costs are a step-function of distance. We focus on locations which are within 2.5 kilometers on one or the other side of a distance threshold where the postage cost changes, which returns 98 observations.[34] It is worth noting that our data include towns in every direction from London, so two locations at a similar distance are not necessarily near each other. Denoting the group of towns around a threshold B as G_B, we estimate:

$$\ln L_{ni} = \beta_0 - \mathbf{1}[n \in G_B] - \eta \ln c_{ni} + \phi \ln \sum_{i' \neq n} P_{i'}^{\beta} MA_{i'}^{\kappa} d_{ni'}^{-\gamma} c_{ni'}^{-\eta} + \beta' X_n + \varepsilon_n \tag{6}$$

which again directly follows from eq. 3. Here, $\mathbf{1}[n \in G_B]$ is a set of group dummies, and X_n include $n-$specific controls. The most important control are population and market access, but we also explore the impact of other controls. The group bins flexibly control for different distances to London across sets of towns. Within each group, locations are at a similar distance to London, but because they fall on different sides of a cost step, post towns that are slightly further away experience a discrete jump in their postage costs. This second equation recovers an estimate for the elasticity of letter flows to postage costs η while flexibly controlling for travel costs using the group fixed effects.

To estimate eq. 5 and 6 we need to address three challenges.

First, we need to construct the market access term MA_n, which is a function of the unknown elasticity θ. In our main analysis, we set $\theta = 6$, which is the central value for the range of elasticities that the trade literature has estimated or used in the past: these values have varied between $\theta = 1$ (Harris, 1954), $\theta = 3.6$ (Bernard *et al.*, 2003), $\theta = 4.14$ (Simonovska & Waugh, 2014), $\theta = 4.87$ (Eaton *et al.*, 2011) , θ

[34] This distance is the largest for which no post town is simultaneously to the right of one threshold and to the left of the next.

between 2.7 and 6.7 (Costinot *et al.*, 2011), θ=8.28, with upper range extending to $\theta = 12.86$ (Eaton & Kortum, 2002). In robustness exercises for our patent analysis, we replicate our main analysis over a grid of integers between $\theta = 1$ and $\theta = 13$.

Second, since we cannot use origin fixed effects, consistent estimation of eq. 5 and 6 mandates that we construct a value for LMA from the data. This fact requires us to determine the set of other locations i' with which residents in post town n could plausibly communicate in 1838. In our data, some registration districts do not have post towns, which presumably strongly limited two-way communication flows. This also means that we do not have direct measures of the postage cost between any post town in our data and residents of those districts.[35] In our main analysis, we then construct LMA always assuming that no communication was possible from any post town to those districts, and vice-versa. In robustness exercises, we make some additional assumptions to assign a postage cost even to districts with no post towns and replicate our patent analysis.

Third, and relatedly, the need to construct and explicitly control for LMA also requires to have a value for the elasticities β, κ, γ and η that enter the summations in eq. 5 and 6. We then make use of an iterative procedure. Given any set of current guesses $\hat{\beta}$, $\hat{\kappa}$, $\hat{\gamma}$ and $\hat{\eta}$, we can construct a proxy for LMA_n, denote it $\widehat{LMA}$, as

$$\ln \widehat{LMA(\hat{\beta}, \hat{\kappa}}, \hat{\gamma}, \hat{\eta}) = \ln \sum_{i' \neq n} P_{i'}^{\hat{\beta}} MA_{i'}^{\hat{\kappa}} d_{ni'}^{-\hat{\gamma}} c_{ni'}^{-\hat{\eta}} \tag{7}$$

This proxy is the direct empirical counterpart to eq. 2. We can then estimate eq. 5, and recover new guesses for β, κ, γ; estimate eq. 6 and obtain an updated guess for η; and obtain a new estimate for $\widehat{LMA}$. We iterate over this procedure until the elasticities estimated in eq. 5 and 6 are the same as those used to construct $\widehat{LMA}$.[36]

The results of the estimation of the first equation are reported in Table 2. Column 1 looks at the relationship between letter flows and travel costs, controlling only for region and postage costs fixed effects; in Column 2 we add in controls for a location's population and market access; we finally add control for letter market access in Column 3. In this preferred specification, as expected, higher travel costs between one location and London significantly reduce the letters flow to London. Locations with a larger population, or larger market access, send more letters to London. On the other hand, origins with greater letter market access, indicating that they have more nearby population centers to communicate with, other than London, send fewer letters to London (controlling for their own size, market access, and travel costs to London). Note that although we have not imposed it, the coefficient on LMA is not statistically different from -1, as eq. 2 would predict.

The estimates for the second equation are reported in Table 3. We estimate postal cost elasticities

[35] These districts account for 16% of the population of England and Wales.

[36] We start with an initial proxy for LMA which sets $\hat{\beta} = 1$, $\hat{\kappa} = 0$, and $\hat{\gamma} = \hat{\eta} = -1$. We have experimented varying our initial conditions. Our convergence criterion stops the algorithm when all coefficients differ from the previous estimates by at most 0.1%.

Table 2: Impact of distance and population on letter flows, controlling for postage cost

	DV: Log Letter Flows to London		
	1	2	3
Ln Travel Costs, 1838	-0.074	0.259	-1.288***
	(0.182)	(0.184)	(0.460)
Ln Population		1.027***	1.094***
		(0.090)	(0.087)
Ln MA		0.108***	0.222***
		(0.024)	(0.027)
Ln $\widehat{LMA}$			-1.469***
			(0.351)
Postage cost FE	Y	Y	Y
Region FE	Y	Y	Y
N	579	579	579
R^2	0.14	0.39	0.44

Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

ranging between -0.98 and -1.54; these estimates become more precise as controls are added. Column 1 only includes group threshold and region fixed effects. Column 2 controls for local population and market access. In Column 3 we include two controls that were not included in the previous Table 2. The first of these is the number of "privileged" letters sent from a location to London, i.e., letters that could be posted without charge (e.g., because a sender had franking privileges). This flow will reflect in part the impact of travel costs, but it will also capture other local features such as the presence of more MPs or government officials, that may have affected letter flows. Since this variable will reflect in part the role of travel costs, we do not want to include it in the previous table, where it is likely to be a bad control, but here it can help us control for additional factors other than cost that impact letter flows. The second new variable, newspaper flows, is included for a similar reason. Like privileged letter flows, newspaper flows were not subject to the same postage scheme as regular letters, so including this control helps us deal with unobserved local factors other than cost that may have influenced letter flows. Columns 4 controls for LMA. In our preferred estimation, the elasticity of the volume of letters to postage costs is estimated at about -1.5.

Our two-part procedure allows us to recover the crucial elasticity of letter flows to postage costs leveraging the discontinuity of those costs around particular distance thresholds, at the price of relying

Table 3: Impact of cost on letter flows

	DV: Log Letter Flows to London			
	1	2	3	4
Ln Postage Cost	-0.908	-1.250	-1.155	-1.541*
	(1.716)	(1.512)	(0.865)	(0.890)
Ln Population		Y	Y	Y
Ln MA		Y	Y	Y
Ln Privileged Letters			Y	Y
Ln Newspaper Flows			Y	Y
Ln $\widehat{LMA}$				Y
Group threshold FE	Y	Y	Y	Y
Region FE	Y	Y	Y	Y
N	98	98	98	98
R^2	0.29	0.45	0.82	0.83

Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

on a small sample of origins. In our patent analysis, we will also perform robustness where our elasticities are recovered with a single-equation approach that estimates eq. 3 directly.

In summary, this analysis yields two main findings. First, it is clear (though not surprising) that postage costs reduce letter flows, and that this effect operates independently of travel costs, which also significantly reduce letter flows. Second, this analysis provides specific estimates of the elasticity of letter flows with respect to postage and travel costs, population, and market access. In our patents analysis, these elasticities are inputs in the construction of the size of the treatment that each location receives from the reform.

5 Main analysis

5.1 Citation data analysis

In this section, we analyze the impact of the reduction in communication costs generated by the postal reform on the exchange of scientific knowledge, as reflected in citations in scientific articles. Since these data provide a bilateral measure of knowledge flows between pairs of locations, we take advantage of standard approaches developed by trade economists for the study of how trade frictions affect trade flows, an analog to our interest in how communication costs affect knowledge flows.

Using n subscripts to denote origin locations (authors) and i subscripts to denote destinations (cited scientists), our primary regression specification is,

$$CITE_{nit} = \beta COST_{ni} \times POST_t + \xi_{nt} + \gamma_{it} + \psi_{ni} + \epsilon_{nit} \quad (8)$$

In this equation, $CITE_{nit}$ is the sum of citations in period t that originate from authors in post town n and are directed towards scientists in post town i. $COST_{ni}$ is the cost of sending a letter from location n to i in the pre-reform period, $POST_t$ is an indicator for the post-reform period, and ξ_{nt}, γ_{it}, and ψ_{ni} are, respectively, origin-period, destination-period, and dyad fixed effects. The inclusion of origin-time and destination-time fixed effects is natural given our data generating process, in which the publication of a single article by an author in an origin location can potentially include citations to a number of destination locations. The inclusion of origin-time and destination-time fixed effects soaks up the impact of simply having an article appear, focusing attention instead on whether there are changes in the extent to which articles cite scientists in locations that were, in terms of postal cost, more distant.

The origin-time and destination-time fixed effects also absorb variation induced by three other factors. The first of these is the service provision of local penny posts, which served local areas around many post towns. The second is the presence of individuals with franking privileges, who were not treated by the reform because they could send letters at zero cost in the pre-reform period, at an origin or destination location. The third factor is the local impact of any changes in international postage rates. More generally, these fixed effects capture any origin trends and destination trends in unobserved factors that might have affected the number of articles or citations produced. The inclusion of dyad fixed effects is also important, since it will absorb fixed pair features including, most importantly, the distance between any two locations.

We exclude citations where both scientists were located within the historic County of London from our citation analysis. Within this area, the close proximity of scientists meant that they could likely communicate by visiting one another, and so we would not expect their bilateral communications to be affected by the reform in the same way as scientists living in other locations. This also makes our citation analysis consistent with our patent analysis, where we will also exclude London for reasons explained later.

Recall from the data section that the unit of analysis in our citation data analysis is the post town. With over 600 post towns, our data set includes a large number of potential origin-destination pairs. Compared to this large matrix, the actual number of citations is relatively small, and the majority of connected pairs are connected by only one citation. Summary statistics are available in Appendix B.3. Given the sparsity of the citations data at the annual level, we collapse the data into one pre-reform period spanning 1830-39 and one post-reform period covering 1840-49.[37] We then estimate our specification using

[37] Bertrand *et al.* (2004) also show that collapsing the data into two periods provides an effective way of addressing potential serial correlation concerns.

PPML. Note that the dyad fixed effects in eq. 8 will cause origin-destination pairs with equal citation counts in both the pre- and the post-reform periods to drop out of the analysis.[38]

Results of our analysis of the citation data are presented in Table 4. In Column 1, we present results with origin-time and destination-time fixed effects, but omitting dyad fixed effects. This allows us to separately estimate the impact of the distance-based postage cost on citations in the pre-reform period. The coefficient on the "Log Cost" variable tells us that citations are substantially lower between locations with higher bilateral postage costs in the pre-reform period. The coefficient on the "Log Cost × Post-reform" period indicates that locations with higher bilateral postage costs experienced a substantial relative increase in citations in the post-reform period. In terms of magnitude, the estimated increase in citations in the pre-reform period is large enough to offset around 70% of the penalty imposed by higher costs in the pre-reform period (with the difference likely due to other impacts of distance on citations that were independent of cost). Put another way, our results indicate that 70% of the decay of knowledge exchange with distance-based postage cost in the pre-reform period, as reflected in citations, disappears as a result of the reform.

In Column 2, we include dyad fixed effects. These absorb the impact of pre-reform costs as well as any other time-invariant factor related to bilateral distance. However, with the inclusion of this large set of additional fixed effects our results are only strengthened. Note that the sample size falls in this regression because any dyad without any citations will drop out. In Column 3, we go even further, by including directed dyad fixed effects. Even in this very stringent specification we still observe a strong impact of postal costs on bilateral citations. The coefficients are larger when using directed dyad fixed effects, but the difference is not statistically distinguishable from Column 2.

In Columns 4-6, we provide similar results but dropping Scotland from the analysis. We do this only for consistency with our patent analysis, where we only have outcome data for England and Wales. These results are slightly larger in magnitude, but also somewhat noisier due to the smaller sample size.

The estimated coefficients in Table 4 suggest that the elasticity of citations with respect to a reduction in postage cost is, in our preferred specifications including dyad fixed effects (Columns 2 and 3), between 0.895 and 1.235 in the full sample. One way to interpret this is relative to a one standard deviation reduction in the log cost, 0.28 in the full sample, which would imply an additional 0.25 to 0.35 citations between a dyad, relative to a sample mean of 0.19 citations with a standard deviation of 1.32 citations in the pre-period estimation sample. Or put another way, the change in log cost between our average-cost and lowest-cost dyads in the sample was about 1.47 log points, so lowering the postage cost for the average cost dyad to that of the lowest would lead us to expect around 1.3-1.8 additional citations over the ten-year post-reform period. If we focus instead only on England and Wales (where we see larger effects but smaller differences in bilateral costs) the estimated impact of a move from the average to the lowest-cost dyads in the sample is an additional 2.5-2.9 bilateral citations across a decade. Of course, these figures only apply to locations where a cited scientist was present sometime during our sample period.

[38] All regressions in this section report the number of observations effectively used in the estimation.

Table 4: Citations analysis results

	DV: Number of citations between an origin-destination pair in a period					
Sample:	Full			Excluding Scotland		
	1	2	3	4	5	6
Ln cost	-1.198***			-1.644***		
	(0.265)			(0.431)		
Ln Cost x Post Reform	0.834**	0.895***	1.235***	1.412***	2.118**	1.873*
	(0.341)	(0.331)	(0.315)	(0.542)	(1.023)	(1.129)
Citing Loc. x Period FE	Y	Y	Y	Y	Y	Y
Cited Loc. x Period FE	Y	Y	Y	Y	Y	Y
Dyad FE	N	Y	N	N	Y	N
Directed Dyad FE	N	N	Y	N	N	Y
N	3,988	345	198	2,358	187	108

Robust standard errors are presented in parenthesis. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 4 presents results with robust standard errors, but we have also estimated results with standard errors clustered at the origin-decade level, which may better account for the fact that multiple citations come from the same article. However, this clustering makes essentially no difference for the standard errors, as shown in Appendix Table B5, which is why we have chosen to focus on robust standard errors in the main text.

To summarize, our citation results show clear evidence that scientific citations increased between location pairs that experienced a greater reduction in bilateral postage costs as a result of the introduction of the uniform penny post. These results are found using a fairly strong analysis strategy that accounts for location-time and location-pair fixed effects. While the economic importance of the additional citations is difficult to assess, the fact that citations respond strongly to reduced communication costs indicates that reducing the cost of long-distance communication played a meaningful role in facilitating knowledge exchange between scientists. Next, we consider whether lower communication costs also facilitated the development of new technologies.

5.2 Patent data analysis

Analyzing patent data provides a useful complement to our analysis of scientific citations. The new technologies represented by new patent filings provide a second margin along which we can assess the impact of lower communication costs, and one that is more directly related to productivity and economic growth.

The nature of the patent data implies that this analysis will differ in important ways from our analysis of citation data. Most importantly, the patent data represent a location-level rather than a bilateral pair-level outcome. This feature has two consequences. First, it means that we need to construct a location-level measure of treatment. Second, it means that we will have to control for factors, such as the population of a location or its market access, instead of being able to include time-locations fixed effects that absorb these concerns.

To construct a location-level measure of the impact of the postal reform, we use an approach similar to the market access measure of Donaldson & Hornbeck (2016). As we have argued in Section 4, one can think of the letter market access as a measure that captures the availability of communication opportunities for residents of a particular place. This measure is a function of postage and travel costs to surrounding destinations, and of those destinations' attractiveness as receivers of letters. The postal reform only changes the postage cost. Consequently, the change in letter market access for location n implied by the reform – a measure of the treatment received by location n – can be defined as

$$\Delta LMA_n = \underbrace{\ln\left(\sum_{i'\neq n} P_{i'}^{\beta} MA_{i'}^{\kappa} d_{ni'}^{-\gamma} 1^{-\eta}\right)}_{LMA_{n,after}} - \underbrace{\ln\left(\sum_{i'\neq n} P_{i'}^{\beta} MA_{i'}^{\kappa} d_{ni'}^{-\gamma} c_{ni'}^{-\eta}\right)}_{LMA_{n,before}} \tag{9}$$

This value can now be computed, since P_i and MA_i are the population and market access of n, d_{ni} is the bilateral travel cost over the waterways, rail and turnpike networks, and c_{ni} the postage cost between locations n and i in the pre-reform period, all observed; and the elasticities β, κ, γ and η have been estimated in Section 4.[39]

Equipped with a measure of the treatment size for a location, ΔLMA_n, we can now analyze the impact of the reform on new patents using standard panel data methods. The unit of observation in our patent data is the post town and, as in the citation analysis, we collapse the data into one pre-reform period (1830-39) and one post-reform period (1840-49). Since, even in the collapsed data, patents are sparse at the location-by-period level for many smaller locations, we estimate results using PPML. London is excluded entirely from our patent analysis because it produced far more patents than any other location and likely had a much different innovation environment (and one that was likely to be much less affected by the fall in postage costs). Our patent analysis regression specification is,

$$PAT_{nt} = \beta_0 POST_t + \beta_1 \Delta LMA_n \times POST_t + X_{nt}\Gamma_0 + \Gamma_1 + \epsilon_{nt} \tag{10}$$

In this equation, where PAT_{nt} is the number of patents associated with post town n in period t, ΔLMA_n

[39]Our preferred specification uses elasticities from Tables 2 and 3. We explore the robustness of our results to alternative methods of calculating treatment, market access, and standard errors in detail in Appendix B.6.

is the change in letter market access defined above, $POST_t$ is an indicator for the post-reform period, X_{nt} is a set of control variables, and Γ_1 is a set of post town fixed effects. In our preferred specification, we include controls for local population and market access from eq. 4 with $\theta = 6$.[40] Our preferred specification also includes an interaction between $POST_t$ and the distance of the post town to London. Even though London's influence will already be reflected as a part of the market access, this control provides additional flexibility that reflects the fact that London may have influence beyond that reflected in its population alone, because so many of the country's institutions, including those such as the Royal Society or the patent office, were located there. Our preferred specification also includes county-by-decade fixed effects, which can help account for the fact that Britain at this time was composed of a set of regional economies that were often on very different growth trajectories.

Regression results for our patent data analysis are presented in Table 5. In all regressions, we cluster the standard errors at the registration district $\times$ decade level because population information only comes at this broader level of aggregation, and some post towns are located in the same district. Column 1 presents the most parsimonious specification, without controls for population, market access, or other time-varying factors. We estimate a positive association between the change in letter market access induced by the treatment and the number of patents, but this relationship is not statistically significant.

In Column 2, we add in controls for population, market access, and distance to London interacted with $POST_t$. Given our empirical setting, these seem like obvious controls to include. Once these controls are included, we begin to observe a stronger and statistically significant relationship between the change in letter market access and patents. Of these controls, changes in local population seem to have the most substantial effect on patenting. Changes in market access do not appear to increase patenting in the time frame that we study (note that initial differences in market access may affect patenting, but this is absorbed by the post town fixed effects). Similarly, we do not observe strong time-varying effects associated with a location's distance to London.

In Column 3, we add county-by-decade fixed effects. These fixed effects help us account for the fact that the Industrial Revolution was heavily concentrated in certain regions of the country, such as Lancashire, the West Riding, and the Midlands, while other areas were left behind. The inclusion of county-by-decade controls can help account for these substantial regional differences. This within-county comparison means we are identifying our estimates off of less variation, in terms of the change in letter market access, but it also means we are comparing locations exposed to more similar economic conditions. With these controls included, we find an even clearer relationship between the change in letter market access and the number of patents in a location. However, we can also see, from the results in Column 2, that the inclusion of these fixed effects is not vital for generating significant results.[41]

[40] Note that market access is time-varying because the transportation network is changing over time.

[41] The increase in the coefficient between Columns 2 and 3 is not a consequence of the slightly different estimation sample that results after introducing the fixed effects; re-estimating Column 2 on the estimation sample in Column 3 yields essentially unchanged estimates.

In terms of magnitude, our preferred estimates in Columns 2-3 suggest that a location that experienced a one standard deviation greater improvement in letter market access as a result of the reform would have produced, on average, an additional 0.075 to 0.114 patents during the post reform period. This is equivalent to a 2.2% to 3.3% increase in the average number of patents produced by locations in the pre-reform period. These magnitudes indicate that reducing communication costs can have a meaningful effect on the rate of innovation in an economy.

Table 5: Patent data analysis results

	DV: Number of patents		
	1	2	3
Post 1840	0.591***	0.526	
	(0.043)	(0.413)	
ΔLMA_n × Post 1840	0.029	0.075*	0.114*
	(0.048)	(0.045)	(0.063)
Ln Population		1.938***	2.157***
		(0.590)	(0.729)
Ln MA		-0.014	-0.027
		(0.128)	(0.146)
Distance to London × Post 1840		-0.048	-0.129
		(0.079)	(0.215)
Post Town FE	Y	Y	Y
County × Decade FE	N	N	Y
N	698	698	688

ΔLMA and *Ln* MA have been standardized. Standard errors clustered by registration district × period in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

In Table 6, we present some robustness results. In Column 1, we follow the same analysis approach used in our preferred specification (Column 3 of Table 5), but we use bootstrapping at the county level to calculate the standard errors.[42] This approach may better accommodate the fact that the ΔLMA term used in our analysis is calculated using parameters estimated from the letter flows regressions in Section 4. However, we can see that using bootstrapped standard errors does not have a substantial influence on our results.

[42]See Appendix B.6 for more details.

In Columns 2 and 3, we present analysis based on alternative approaches to calculating the parameters used in the ΔLMA term. Recall that in our estimation of letter market access in Section 4, our data includes only communication destinations with post towns. This excludes the population of some rural districts with no post towns from the letter market access of a given post town, even though that population can still influence market access. This is a reasonable choice since the absence of a post town would have likely limited communication flows. However, for the results in Column 2, we add back in these districts by connecting them to the nearest post town and then assuming a one-penny additional cost for this connection.[43] We then re-estimate the letters-level regression, the treatment size, and our preferred specification. This approach gives us larger and more statistically significant estimates of the effect of changing communication costs on patenting.

In Column 3, we revert to the data set used in our baseline letter market access specification, but take an alternative approach to estimation. Instead of the two-part estimation approach used in our preferred results, in this specification we use parameters obtained from a single regression that includes both travel cost and postage cost.[44] We can see that using this alternative approach to estimating the parameters used in ΔLMA gives results that are similar, though slightly larger and slightly noisier, than those produced by our preferred approach.

In Appendix B.6, we report some additional robustness exercises. In particular, Table B7 replicates our preferred specification of Column 3 in Table 5 and all columns of Table 6 using a wide range of different values of the trade elasticity, θ. We let this parameter vary over a grid of integers between 1 and 13: as discussed in Section 4, this range encompasses most of the estimates found or used in the literature. Note that the market access of a location, which is a function of θ, appears in the estimation of the letters-level regression in Section 4. Hence, for each value of θ we first re-estimate the elasticities of letters flow to population, market access, travel and postage costs; we then recompute the treatment size ΔLMA_n; and finally, report our results on the effect of the reform. We find that our results are not sensitive to the value of θ used in the analysis.

The main message from these robustness exercises is that we find consistent evidence that the reduction in communication costs generated by the reform, which resulted in differential increases in letter market access across locations, was associated with an increase in patenting. Our estimated coefficients do vary across specifications, ranging from 0.11 to 0.44. This reflects in part the fact that we are working with a relatively modest sample size in both our patent regressions and in the letter flows regressions that generate the parameters used to construct our key treatment variable. Relative to the range of effects that we estimate, our preferred specifications in Table 5 deliver results that are on the conservative side.

[43] This assumption would be consistent with the working of the postal system if there were "local penny posts" that would deliver postage to the main post offices, although it likely introduces significant noise in the calculation of postage costs.

[44] We prefer our two-equations approach to this alternative because the correlation between travel costs and postage costs, and the fact that the first of these is likely to be measured with more error than the second, makes it challenging to clearly separate the effect of postage cost from distance. Our preferred approach exploits instead the sharp increase in postage costs around distance thresholds. Appendix B.6 contains more details on this alternative procedure.

Table 6: Patent data analysis results: robustness

	DV: Number of patents		
	Baseline approach with bootstrap SEs	Alternative LMA: communication includes all districts	Alternative LMA: single-equation estimation
	1	2	3
ΔLMA_n × Post 1840	0.114*	0.389***	0.271*
	(0.062)	(0.144)	(0.159)
Ln Population	2.157**	1.959***	2.311***
	(0.859)	(0.679)	(0.878)
Ln MA	-0.027	-0.106	0.001
	(0.222)	(0.145)	(0.138)
Distance to London × Post 1840	-0.129	-0.604**	-0.507
	(0.363)	(0.269)	(0.331)
Post Town FE	Y	Y	Y
County × Decade FE	Y	Y	Y
N	688	688	688

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. In all three columns, ΔLMA and $Ln\ MA$ have been standardized. In Column 1, we follow our preferred analysis approach from Column 3 of Table 5, but using bootstrapped standard errors (columns 2-3 revert to standard errors clustered by registration district times decade). In Columns 2-3, we present results in which we calculate LMA using parameters obtained using different approaches to calculating letter market access. In Column 2, we include additional districts that do not have a post town in our letter market access regressions. In Column 3, we use parameters obtained from a single letter market access regression that includes both travel cost and postage cost terms, rather than the two-part approach used in our preferred specification.

The evidence in this subsection indicates that the reduction in communication costs induced significant increases in patenting activity in more exposed relative to less exposed post towns. Taken together, the findings in this section are consistent with a central role of reduction in communication costs in enhancing the circulation of scientific knowledge and the creation of new ideas.

One lingering question to emerge from the preceding analysis has to do with the extent to which the technological developments reflected in the patent data may have been linked to the basic science reflected in the citation data. This issue – the link between basic science and technological development during the Industrial Revolution – is the subject of a long and ongoing debate among economic historians (see, e.g., Landes (1969), Rosenberg (1974), Mokyr (2002), Khan (2018), Jacob (2014), and Kelly & Ó Gráda (2020)). Establishing the direct link between scientific knowledge and technological development has often proven elusive, mainly because it is often difficult to establish links between basic science and new technologies. However, there seems little doubt that over a sufficiently long time horizon, technological development depends crucially on the development of basic scientific knowledge.

6 Conclusions

Economists have long suspected that the changes in the cost of exchanging knowledge are likely to influence the rate at which useful new ideas are developed. Mokyr (2005b), for example, writes that, "access to useful knowledge created the opportunities to recombine its components to create new forms that would expand the volume of knowledge at an even faster rate." By taking advantage of the large and spatially varied reduction in communication costs resulting from the introduction of the Uniform Penny Post, our study provides more direct evidence on the impact of communication costs on innovation rates and the exchange of scientific knowledge than has heretofore been available. Our findings confirm the long-held belief that knowledge flows matter for science and innovation, and help place the extensive theoretical literature embodying these ideas on a more solid empirical foundation.

Establishing a link between communication costs and innovation has particular significance for some theories of endogenous growth. For example, in discussing the sources of sustained increase in the standards of living in modern capitalist economies, Lucas (2009) argues, "What is central, I believe, is the fact that the industrial revolution involved the emergence (or rapid expansion) of a class of educated people, thousands—now many millions—of people who spend entire careers exchanging ideas, solving work-related problems, generating new knowledge." Our results suggest that the "generation" of ideas is intimately linked to the "exchange" of ideas, and they speak to a large growth literature which assumes ideas diffusion as an engine of economic growth. Our findings are also significant for our understanding of cities. As Davis & Dingel (2019) write, "Leading empiricists and theorists of cities have recently argued that the generation and exchange of ideas must play a more central role in the analysis of cities." To the extent that proximity reduces communication costs, our results suggest that cities are accelerating the generation and circulation of knowledge. This may help explain why so much innovation takes place in cities.

Our results also contribute to our understanding of innovation in Britain during the Industrial Revolution. Joel Mokyr has argued that "The true miracle is not that the Industrial Revolution happened, but that it did not peter out like so many earlier waves of innovation" (Mokyr, 2004). Our findings suggest that institutional reforms may have played an important role in sustaining technological progress during this crucial period of economic history.

References

Abhay, Aneja, & Xu, Guo. 2022 (3). *Strengthening State Capacity: Postal Reform and Innovation during the Gilded Age*. NBER Working Paper No. 29852.

Acemoglu, Daron, Moscona, Jacob, & Robinson, James A. 2016. State Capacity and American Technology: Evidence from the Nineteenth Century. *American Economic Review*, **106**(5), 61–67.

Aghion, Philippe, & Howitt, Peter. 1996. Research and Development in the Growth Process. *Journal of Economic Growth*, **1**(1), 49–73.

Agrawal, Ajay, & Goldfarb, Avi. 2008. Restructuring research: Communication costs and the democratization of university innovation. *American Economic Review*, **98**(4), 1578–90.

Agrawal, Ajay, Galasso, Alberto, & Oettl, Alexander. 2017. Roads and innovation. *Review of Economics and Statistics*, **99**(3), 417–434.

Akcigit, Ufuk, Hanley, Douglas, & Serrano-Velarde, Nicolas. 2021. Back to Basics: Basic Research Spillovers, Innovation Policy, and Growth. *Review of Economic Studies*, **88**, 1–43.

Aker, Jenny C. 2010. Information from markets near and far: Mobile phones and agricultural markets in Niger. *American Economic Journal: Applied Economics*, **2**(3), 46–59.

Allen, Treb. 2014. Information Frictions in Trade. *Econometrica*, **82**(6), 2041–2083.

Alvarez-Palau, Eduard J, Dunn, Oliver, Bogart, Dan, Satchell, Max, & Shaw-Taylor, Leigh. 2019. *Historic ports and coastal sailing routes in England and Wales, 1540-1914. [Data Collection]*. Colchester, Essex (UK): UK Data Service - Reshare. 10.5255/UKDA-SN-853711.

Audretsch, DB, & Feldman, MP. 1996. R&D spillovers and the geography of innovation and production. *American Economics Review*, **86**(3), 630–640.

Basire, James. 1838. *Sketch of the Post Office Map February 1838*. Classical Images: Aniquarian Maps, Prints & Books, https://perma.cc/B9UQ-AL2D.

Belenzon, Sharon, & Schankerman, Mark. 2013. Spreading the word: Geography, policy, and knowledge spillovers. *Review of Economics and Statistics*, **95**(3), 884–903.

Bernard, Andrew B., Eaton, Jonathan, Jensen, J. Bradford, & Kortum, Samuel. 2003. Plants and Productivity in International Trade. *American Economic Review*, **93**(4), 1268–1290.

Bertrand, Marianne, Duflo, Esther, & Mullainathan, Sendhil. 2004. How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, **119**(1), pp. 249–275.

Bogart, Dan, You, Xuesheng, Alvarez-Palau, Eduard J., Satchell, Max, & Shaw-Taylor, Leigh. 2022. Railways, divergence, and structural change in 19th century England and Wales. *Journal of Urban Economics*, **128**, 103390.

Bottomley, Sean. 2014. *The British Patent System during the Industrial Revolution 1700–1852*. Cambridge University Press.

Boudreau, Kevin J, Brady, Tom, Ganguli, Ina, Gaule, Patrick, Guinan, Eva, Hollenberg, Anthony, & Lakhani, Karim R. 2017. A field experiment on search costs and the formation of scientific collaborations. *Review of Economics and Statistics*, **99**(4), 565–576.

British Almanac. 1828-1875. *The British Almanac of the Society for the Diffusion of Useful Knowledge. 1828:* Baldwin & Cradlock, *1829-1869:* Knight & Co, *1870-1875:* The Company of Stationers, London.

Buera, Francisco J., & Oberfield, Ezra. 2020. The Global Diffusion of Ideas. *Econometrica*, **88 (1)**, 83–114.

Carlino, Gerald A, Carr, Jake, Hunt, Robert M, Smith, Tony E, *et al.* 2012. *The agglomeration of R & D labs*. Tech. rept. Federal Reserve Bank of Philadelphia.

Cary, John. 1828. *Cary's new itinerary: or, An accurate delineation of the great roads, both direct and cross throughout England and Wales; with many of the principal roads in Scotland.* G. & J. Cary, London.

Catalini, Christian, Fons-Rosen, Christian, & Gaulé, Patrick. 2020. How Do Travel Costs Shape Collaboration? *Management Science*, **66**(8), 3340–3360.

Chapman, Allan. 2011. Airy, Sir George Biddell. *In: Oxford Dictionary of National Biography*. Cambridge University Press, accessed 01/14/19.

Clifton, Gloria. 1995. *Directory of British Scientific Instrument Makers 1550-1851*. Philip Wilson Publishers Ltd, London, UK.

Coase, R. H. 1939. Rowland Hill and the Penny Post. *Economica*, **6 (24)**, 423–435.

Cobb, M.H. 2006. *The Railways of Great Britain: a Historical Atlas at the Scale of 1 Inch to 1 Mile. 2 vols*. Shepperton: Allen.

Colquhoun, Patrick. 1806. *A Treatise on Indigence*. Hatchard.

Costinot, Arnaud, Donaldson, Dave, & Komunjer, Ivana. 2011. What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas. *The Review of Economic Studies*, **79**(2), 581–608.

Davis, Donald R., & Dingel, Jonathan I. 2019. A Spatial Knowledge Economy. *American Economic Review*, **109**(1), 153–70.

Donaldson, Dave. 2018. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, **108**(4-5), 899–934.

Donaldson, Dave, & Hornbeck, Richard. 2016. Railroads and American Economic Growth: A "Market Access" Approach. *Quarterly Journal of Economics*, **131**(2), 799–858.

Duranton, G., & Puga, D. 2004. Handbook of Regional and Urban Economics. North-Holland.

Dutton, Harold I. 1984. *The Patent System and Inventive Activity During the Industrial Revolution, 1750-1852*. Manchester University Press.

Eaton, Jonathan, & Kortum, Samuel. 2002. Technology, Geography, and Trade. *Econometrica*, **70**(5), 1741–1779.

Eaton, Jonathan, Kortum, Samuel, & Kramarz, Francis. 2011. An anatomy of international trade: Evidence from French firms. *Econometrica*, **79**(5), 1453–1498.

Fava-Verde, JF. 2018. Victorian Telegrams: The Early Development of the Telegraphic Dispatch and Its Interplay with the Letter Post. *Notes and Records*, **72**, 275–292.

Feigenbaum, J, & Rotemberg, M. 2014. *Communication and Manufacturing: Evidence from the Expansion of Postal Services*. Mimeo.

Fyfe et al., A. 2017. Key Facts Data Sheet for Philosophical Transactions in 1840. *In: Publishing the Philosophical Transactions, 1665-2015*.

General Post Office. 1840. *Post Office Regulations*. Handbill, January 7th, 1840, J Hartnell.

General Post Office. 1855. *First Report of the Postmaster General on the Post Office*. George E. Eyre and William Spottiswoode.

General Post Office. 1856. *Second Report of the Postmaster General on the Post Office*. George E. Eyre and William Spottiswoode.

Goyal, Aparajita. 2010. Information, direct access to farmers, and rural market performance in central India. *American Economic Journal: Applied Economics*, **2**(3), 22–45.

Griliches, Zvi. 1986. Productivity, R and D, and Basic Research at the Firm Level in the 1970s. *American Economic Review*, **76**(1), 141–154.

Harris, Chauncy D. 1954. The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers*, **44**, 315–348.

Hemmeon, J. C. 1912. *The History of the British Post Office*. Harvard Economic Studies Vol. VII, Published under the Direction of the Department of Economics.

Hill, Rowland. 1840. On the Effect of the New Postage Arrangement upon the Number of Letters. *Journal of the Statistical Society of London*, **3 (1)**, 102–105.

Jacob, Margaret C. 2014. *The First Knowledge Economy*. Cambridge University Press.

Jaffe, AB, Trajtenberg, M, & Henderson, R. 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics* , **108**(3), 577–598.

James, Frank A. J. 1991-2011. *The Correspondence of Michael Faraday*. The Institution of Engineering and Technology, 6 vols.

Jensen, Robert. 2007. The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *The Quarterly Journal of Economics*, **122**(3), 879–924.

Kelly, Morgan, & Ó Gráda, Cormac. 2020. *Connecting the Scientific and industrial Revolutions: The Role of Practical Mathematics*. CEPR Discussion Paper No. 14885.

Khan, B Zorina. 2018. Human Capital, Knowledge and Economic Development: Evidence from the British Industrial Revolution, 1750–1930. *Cliometrica*, **12**(2), 313–341.

Koudijs, Peter. 2014. The boats that did not sail: Asset price volatility in a natural experiment. *Journal of Finance*.

Landes, David S. 1969. *The Unbound Prometheus*. London: Cambridge University Press.

Lucas, Robert E. 1988. On the Mechanics of Economic Development. *Journal of Monetary Economics*, **22**(1), 3–42.

Lucas, Robert E. 2009. Ideas and Growth. *Economica*, **76**(301), 1–19.

MacLeod, Christine. 1988. *Inventing the Industrial Revolution: The English Patent System, 1660–1800*. Cambridge University Press.

MacLeod, Christine, Tann, Jennifer, Andrew, James, & Stein, Jeremy. 2003. Evaluating Inventive Activity: The Cost of Nineteenth-Century UK Patents and the Fallibility of Renewal Data. *The Economic History Review*, **56**(3), pp. 537–562.

Mansfield, Edwin. 1980. Basic Research and Productivity Increase in Manufacturing. *American Economic Review*, **70**(5), 863–873.

Marshall, Alfred. 1890. *Principles of Economics*. New York: Macmillan and Co.

Martí-Henneberg, Jordi, Satchell, Max, You, Xuesheng, Shaw-Taylor, Leigh, & Anthony, Wrigley E. 2017. *England, Wales and Scotland railway stations 1807–1994 shapefile*. https://www.campop.geog.cam.ac.uk/research/occupations/datasets/catalogues/documentation/engwalesscotlandrail_stations1807to1994.pdf.

Mokyr, Joel. 2002. *The Gifts of Athena*. Princeton University Press.

Mokyr, Joel. 2004. Accounting for the Industrial Revolution. *Pages 1–27 of: The Cambridge Economic History of Modern Britain*. Cambridge University Press.

Mokyr, Joel. 2005a. The Intellectual Origins of Modern Economic Growth. *Journal of Economic History*, **65**(2), 285–351.

Mokyr, Joel. 2005b. Long-Term Economic Growth and the History of Technology. *Chap. 17, pages 1113–1180 of:* Aghion, Philippe, & Durlauf, Steven (eds), *Handbook of Economic Growth*, 1 edn., vol. 1, Part B. Elsevier.

Moser, P. 2012. Innovation without Patents: Evidence from World's Fairs. *Journal of Law and Economics*, **55**(1), pp. 43–74.

Murata, Yasusada, Nakajima, Ryo, Okamoto, Ryosuke, & Tamura, Ryuichi. 2014. Localized knowledge spillovers and patent citations: A distance-based approach. *Review of Economics and Statistics*, **96**(5), 967–985.

Nelson, Richard R. 1959. The Simple Economics of Basic Scientific Research. *Journal of Political Economy*, **67**(3), 297–306.

Nuvolari, Alessandro, & Tartari, Valentina. 2011. Bennet Woodcroft and the value of English patents, 1617-1841. *Explorations in Economic History*, **48**(1), 97–115.

O'Byrne, William R. 1849. *A Naval Biographical Dictionary*. John Murray, Albemarle Street, London, UK.

Postage Act. 1795. *An Act for Further Regulating the Sending and Receiving Letters Free from the Duty of Postage*. 35 Geo 3 c.53, 5th May 1795.

Postage Act. 1812. *An Act for Granting Certain Additional Rates of Postage in Great Britain*. 52 Geo 3 c.88, 9th July 1812.

Postage Act. 1839. *An Act for the Further Regulation of the Duties on Postage*. 2 & 3 Vict c.52.

Rogowski, Jon C, Gerring, John, Maguire, Matthew, & Cojocaru, Lee. 2019. Public Infrastructure and Economic Development: Evidence from Postal Systems. *American Journal of Political Science*.

Romer, PM. 1986. Increasing Returns and Long-run Growth. *Journal of Political Economy*, **94**(5), 1002–1037.

Rosenberg, Nathan. 1974. Science, Invention and Economic Growth. *The Economic Journal*, **84**(333), 90–108.

Rosevear, Alan, Satchell, Max, Bogart, Dan, Shaw Taylor, Leigh, Aidt, Toke, & Leon, Gabriel. 2017. *1830 England and Wales navigable waterways shapefile*. https://www.campop.geog.cam.ac.uk/research/occupations/datasets/catalogues/documentation/turnpikeroads16671892.pdf.

Royal Society. 1867-1871. *Catalogue of Scientific Papers (1800-1863)*. Vol. I-V. Eyre & Spottiswoode, London.

Satchell, Max, Shaw-Taylor, Leigh, & Anthony, Wrigley E. 2017. *1830 England and Wales navigable waterways shapefile*. http://www.geog.cam.ac.uk/research/projects/occupations/datasets/documentation.html.

Schaaf, Larry J. 2021. *The Correspondence of William Henry Fox Talbot Project*. Document number: 3433, http://foxtalbot.dmu.ac.uk/project/project.html.

Schmookler, Jacob. 1966. *Invention and Economic Growth*. Harvard University Press.

Select Committee on Postage. 1838. *Second Report from the Select Committee on Postage*. Reports from Committees, Communicated by the Commons to the Lords.

Select Committee on Postage. 1843. *Report from the Select Committee on Postage*. Reports from Committees, Session 2 February - 24 August 1843, vol. VIII.

Simonovska, Ina, & Waugh, Michael. 2014. The elasticity of trade: Estimates and evidence. *Journal of International Economics*, **92**(1), 34–50.

Steinwender, Claudia. 2018. Real Effects of Information Frictions: 'When the States and the Kingdom became United'. *American Economic Review*, **108**(3), 657–696. Mimeo.

Sullivan, Richard. 1989. England's Age of invention: The acceleration of patents and patentable invention during the industrial revolution. *Explorations in Economic History*, **26**(4), 424–452.

Thompson, Peter. 2006. Patent citations and the geography of knowledge spillovers: evidence from inventor-and examiner-added citations. *The Review of Economics and Statistics*, **88**(2), 383–388.

Waldinger, Fabian. 2011. Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany. *The Review of Economic Studies*, **79**(2), 838–861.

Wallace, Robert. 1837-38. *Third report from the Select Committee on Postage; together with an abstract of the evidence, directed by the committee to be appended to the report.* Tech. rept. House of Commons Papers No 708.

Whewell, William. 1836. Researches on the Tides. Sixth Series. On the Results of an Extensive System of Tide Observations Made on the Coasts of Europe and America in June 1835. *Philosophical Transactions of the Royal Society of London*, 289–341.

Woodcroft, Bennet London: Printed by . 1854. *Titles of patents of invention: chronologically arranged from March 2, 1617 (14 James I.) to October 1, 1852 (16 Victoriæ).* London: G.E. Eyre and W. Spottiswoode.

Online Appendix for "A Penny for your Thoughts" (Not for Publication)

W. Walker Hanlon[1]
Northwestern University, CEPR and NBER

Stephan Heblich[2]
University of Toronto and NBER

Ferdinando Monte[3]
Georgetown University, CEPR and NBER

Martin B. Schmitz[4]
Vanderbilt University

May 2022

[1]Department of Economics. Email: whanlon@northwestern.edu.
[2]Munk School of Global Affairs & Public Policy and Dept. of Economics. Email: stephan.heblich@utoronto.ca.
[3]McDonough School of Business. Email: ferdinando.monte@georgetown.edu.
[4]Department of Economics; Email: martin.schmitz@vanderbilt.edu.

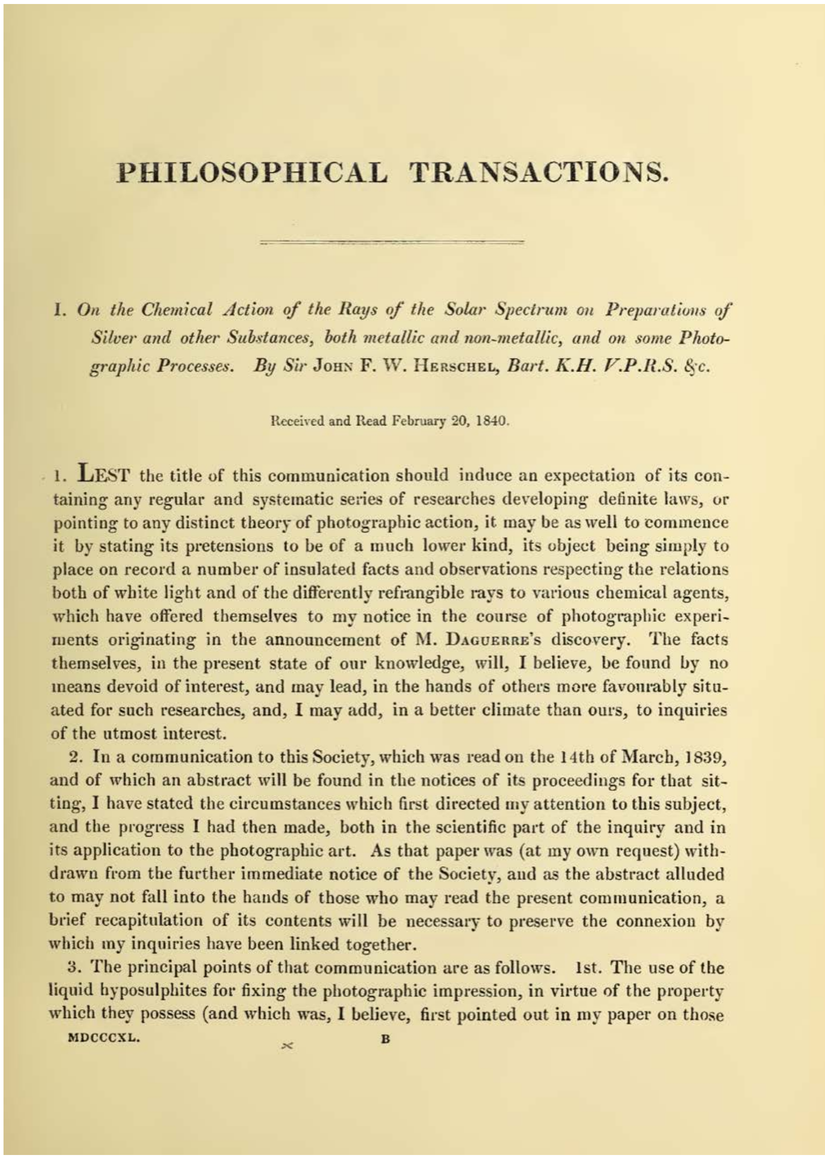
Table of Contents

A Additional Figures

Figure A1: Gross and Net Post Office Revenue, 1830-1849

Note: The figure shows Gross Post Office Revenues (black dashed line), Costs of Management (gray dashed line) and Net Post Office Revenues (black solid line) for the years 1830–1849. Source: House of Commons Papers, Number 421, Vol. 60, Number 707, Vol 95.

Figure A2: An example of citations in the *Philosophical Transactions*



PHILOSOPHICAL TRANSACTIONS.

I. *On the Chemical Action of the Rays of the Solar Spectrum on Preparations of Silver and other Substances, both metallic and non-metallic, and on some Photographic Processes. By Sir* John F. W. Herschel, *Bart. K.H. V.P.R.S. &c.*

Received and Read February 20, 1840.

1. LEST the title of this communication should induce an expectation of its containing any regular and systematic series of researches developing definite laws, or pointing to any distinct theory of photographic action, it may be as well to commence it by stating its pretensions to be of a much lower kind, its object being simply to place on record a number of insulated facts and observations respecting the relations both of white light and of the differently refrangible rays to various chemical agents, which have offered themselves to my notice in the course of photographic experiments originating in the announcement of M. Daguerre's discovery. The facts themselves, in the present state of our knowledge, will, I believe, be found by no means devoid of interest, and may lead, in the hands of others more favourably situated for such researches, and, I may add, in a better climate than ours, to inquiries of the utmost interest.

2. In a communication to this Society, which was read on the 14th of March, 1839, and of which an abstract will be found in the notices of its proceedings for that sitting, I have stated the circumstances which first directed my attention to this subject, and the progress I had then made, both in the scientific part of the inquiry and in its application to the photographic art. As that paper was (at my own request) withdrawn from the further immediate notice of the Society, and as the abstract alluded to may not fall into the hands of those who may read the present communication, a brief recapitulation of its contents will be necessary to preserve the connexion by which my inquiries have been linked together.

3. The principal points of that communication are as follows. 1st. The use of the liquid hyposulphites for fixing the photographic impression, in virtue of the property which they possess (and which was, I believe, first pointed out in my paper on those

MDCCCXL. B

Note: This figure reproduces the first page of the first article that appeared in the 1840 issue of the Philosophical Transactions. "M. DAGUERRE" in the middle of the first paragraph is the first citation in this article. The cited inventor is Monsieur Louis-Jacques-Mandé Daguerre, who invented the first photographic process.

Figure A3: Patents filed during the study period, 1830–1849

Note: The figures shows the number of patents files between 1830–1849. Patent data were compiled by the British Patent Office (Woodcroft, 1854) as part of the 'Titles of Patents of Invention.' These data were digitized by Nuvolari & Tartari (2011).

B Additional Tables

B.1 Calculation of letter volumes

Table B1 presents letter volumes for England & Wales, Scotland, and Great Britain. We extracted the letter volumes for 1839 to 1849 from official publications. We used Select Committee on Postage (1843) p.228 for the year 1839 and General Post Office (1855) p. 56 for the years 1840 to 1849. Although comparable statistics are not available for 1832 to 1838, the fact that postage rates remained constant during these years allows us to back out the letter volumes by calculating the average postage rates for general post letters in England & Wales, Scotland, and Great Britain and dividing these entities' letter-induced annual gross revenues by the corresponding average rate.

Table B1: Letter volumes (in million)

Year	England & Wales	Scotland	Great Britain
1832	61.207	7.495	68.702
1833	60.722	7.366	68.088
1834	63.190	7.439	70.630
1835	64.203	7.581	71.785
1836	67.385	7.938	75.323
1837	67.531	8.011	75.542
1838	61.973	7.152	69.125
1839	65.155	7.959	73.114
1840	132.004	18.554	150.558
1841	154.471	21.235	175.706
1842	163.891	22.216	186.106
1843	173.495	23.473	196.968
1844	189.652	26.502	216.154
1845	214.154	28.669	242.823
1846	235.879	31.135	267.014
1847	253.412	33.261	286.673
1848	260.380	33.563	293.943
1849	267.188	34.747	301.935

Table B2 lists the reports from which we extracted the revenue data. We combine the annual revenue data for the year 1838 with letter volume data for several weeks in the same year to calculate the average pre-reform rate of general post letters in England & Wales, Scotland, and Great Britain (see Select Committee on Postage, 1838, p. 39 et seqq.).

The report that covers the years 1832 to 1834 only gives an account of the net produce in England & Wales, Scotland, and Great Britain in 1833. To approximate the respective gross revenues in 1833, we divide each entity's net produce by the average of the two net-to-gross-produce ratios in 1832 and 1834. Despite an extensive search effort, we were not able to track down a data source that would allow us to extend the letter volume time series further back than 1832.

Table B2: Sources of annual revenue data

Years	Sources
1832-1834	Tables of Revenue, Population, Commerce, etc of the United Kingdom part 4 (1835) for 1834, p. 43
1835-1836	Tables of Revenue, Population, Commerce, etc of the United Kingdom part 6 (1838) for 1836, p. 41
1837	Tables of Revenue, Population, Commerce, etc of the United Kingdom part 7 (1839) for 1837, p. 41
1838-1839	Tables of Revenue, Population, Commerce, etc of the United Kingdom part 9 (1841) for 1839, p. 41

B.2 Summary statistics for letter flow analysis

Table B3 presents summary statistics for the variables used in the letter flows analysis of Section 4. "Letters" is the number of letters sent to London from other post towns around Great Britain in the week starting January 15, 1838. This flow is calculating by subtracting the number of privileged letters from the total number of letters sent as reported in the original source table. "Travel Costs, 1838" is the travel cost between the centroid of the post town and London, computed over waterways, turnpike road, and rail network in 1838, as described in Section 3.4. "Population" is the population of the Registration District where the post town belongs in 1838. The variable $\widehat{LMA}$ is the estimated letter market access of a post town, recovered as described in Section 4 in the main text.

Table B3: Summary statistics for data used in the analysis of letter flows

Variable	Obs	Mean	Std. Dev.	Min	Max
In Levels					
Letters	579	258.399	467.112	1	5660
Travel Costs, 1838	579	187.777	112.462	17.908	531.461
Population	579	24490.52	17921.28	2283.572	210576.2
$\widehat{LMA}$	579	1144.239	1854.07	120.657	21968.98
In Logs					
Ln Letters	579	4.859	1.148	0	8.641
Ln Travel Costs, 1838	579	5.029	.687	2.885	6.276
Ln Population	579	9.944	.543	7.733	12.258
$Ln\widehat{LMA}$	579	6.597	.843	4.793	9.997

B.3 Summary statistics for the citations analysis

Table B4 presents summary statistics for the estimation sample of citations across directed post town dyads in the citation data, by period, and on the log postage cost pre-reform. A directed dyad counts citations from location n towards scientists in location i separately from citations from location i towards scientists in location n. Panel A reports statistics for the full sample. Panel B reports statistics for the sample that excludes citations originating from or towards authors in Scotland. "Pre-reform" refers to the period 1830-1839, and "Post-reform" refers to the period 1840-1849.

Table B4: Summary statistics for the citations data

Variable	Obs	Mean	Std. Dev.	Min	Max
Panel A: full sample					
Citations pre-reform	2117	.187	1.319	0	31
Citations post-reform	1871	.191	1.409	0	44
Log pre-reform bilateral postage cost	2117	2.567	.28	1.099	3.114
Panel B: excluding Scotland					
Citations pre-reform	1219	.233	1.551	0	31
Citations post-reform	1139	.23	1.575	0	44
Log pre-reform bilateral postage cost	1219	2.459	.265	1.099	2.833

B.4 Robustness of citation data analysis

Table B5 replicates the analysis of Table 4 in the main text but clustering the standard errors at the origin-decade level. Columns 1-3 use the full available sample of citations, while Columns 4-6 exclude authors living in, or citations directed to, Scotland. All columns include origin-by-decade and destination-by-decade fixed effects. Columns 2 and 5 include dyad fixed effects, and Columns 3 and 6 include directed dyad fixed effects.

The main message from this Table is that clustering the standard errors maintains the same pattern of significance, while strengthening it in one case.

Table B5: Citations analysis results: clustered standard errors

	DV: Number of citations between an origin-destination pair in a period					
Sample:	Full			Excluding Scotland		
	1	2	3	4	5	6
Log cost	-1.198***			-1.644***		
	(0.225)			(0.463)		
Ln Cost x Post Reform	0.834***	0.895***	1.235***	1.412**	2.118**	1.873*
	(0.282)	(0.316)	(0.218)	(0.609)	(0.908)	(0.980)
Citing Loc. x Period FE	Y	Y	Y	Y	Y	Y
Cited Loc. x Period FE	Y	Y	Y	Y	Y	Y
Dyad FE	N	Y	N	N	Y	N
Directed Dyad FE	N	N	Y	N	N	Y
N	3,988	345	198	2,358	187	108

Standard errors clustered at the origin-decade level are presented in parenthesis. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

B.5 Summary statistics for the patent analysis data

Table B6 presents summary statistics for the estimation sample in the patents analysis exercise."Pre-reform" refers to the period 1830-1839, and "Post-reform" refers to the period 140-1849. "Patents" is the count of patents filed in a location. ΔLMA is the exposure to the reform calculated as described in Section 5.2. "Population" is the population in the Registration District where the post town is located. "Market access" is the travel costs–discounted sum of the population for all Registration Districts in Great Britain, as computed from the perspective of the registration district where the post town is located (eq. 4), using a trade elasticity $\theta = 6$. For the purposes of the patents-level analysis, the measures of exposure to the reform and the log market access have been standardized to provide comparability across specifications.

Table B6: Summary statistics for the patent analysis data

Variable	Obs	Mean	Std. Dev.	Min	Max
Patents, pre-reform	349	3.427	14.105	0	175
Patents, post-reform	349	6.16	25.238	0	324
ΔLMA	349	0	1.001	-3.271	2.262
Ln Population, pre-reform	349	10.051	.541	8.44	12.153
Ln Population, post-reform	349	10.153	.577	8.49	12.394
Ln Market access, pre-reform	349	-.208	.953	-2.56	3.841
Ln Market access, post-reform	349	.208	1.004	-2.149	3.878
Ln Distance to London	349	5.047	.704	2.911	6.102

B.6 Robustness of patent data analysis

Table B7 shows several robustness exercises on the effect of the reform, reporting only the coefficient on $\Delta LMA \times Post$ 1840. Each cell is the result of a separate estimation.

Different rows vary the value of the trade elasticity θ assumed in the market access term 4 that is used in the estimation procedure. As described in Section 4, the trade literature has used or estimated values that broadly range from 1 to 13, which is the interval we focus on in these exercises. For a given value of θ, we recover the elasticities of letter flows to population β, market access κ, travel costs γ and postage costs η. We use these elasticities to compute the treatment received by each location as described in Section 5.2. Finally, we run the PPML regression of eq. 10, controlling for population, market access (at the given value of θ), distance to London interacted with the Post 1840 dummy, and post towns and county-by-decade fixed effects. Table B7 reports the coefficient and standard error on the regressor $\Delta LMA \times Post$ 1840.

Column 1 in this Table corresponds to Column 3 of Table 5, and so it reports robustness on our preferred specification as θ varies.

Column 2 re-estimate our preferred specification across θ bootstrapping the standard errors. We use 1,000 replications clustering at the county level. This clustering is appropriate for two reasons. First, we want to preserve within-county comparisons to let identification be less impacted by spatially correlated unobservable factors. Second, we want to make sure that if one county appears in the pre-reform period, then it also appears in the post-reform period in the same bootstrap replication sample.

Column 3 estimates our model with a different assumption on the set of districts with which residents of a given post town could communicate in 1838. Recall from our discussion in Section 4 that some rural districts do not report a post town. In our main analysis, we have assumed that those districts – for which we cannot in fact compute postage costs – have limited communication with the rest of the country and hence they do not enter the LMA calculation for any post town. For the results in Column 3, we have connected the centroid of these districts to the nearest post town and then assumed an additional one-penny cost for this connection. This assumption would be consistent with the working of the postal system if there were "local penny posts" that would deliver postage to the main post offices, although it likely introduces significant noise in the calculation of postage costs.

Column 4 modifies slightly our estimation procedure to recover the elasticities β, κ, γ, and η. In particular, we use an iterative procedure but estimate directly:

$$\ln L_{ni} = \alpha_0 - \gamma \ln d_{ni} - \eta \ln c_{ni} + \beta \ln P_n + \kappa \ln MA_n - \phi \ln \sum_{i' \neq n} P_{i'}^{\beta} MA_{i'}^{\kappa} d_{ni'}^{-\gamma} c_{ni'}^{-\eta} + \alpha' X_n + \varepsilon_n$$

This expression follows directly from eq. 3 after substituting our assumptions for L_n and $A_{i'}$. Here, α_0 is a constant, d_{ni} is the travel cost between post towns n and i computed over the combined waterways, rail, and turnpike networks in 1838, and c_{ni} is the cost of exchanging letters pre-reform. The terms P_n and MA_n are the 1838 population and market access of the registration district in which the post town is located. The term ε_n captures the sum of classical measurement error in the regressors. The X_n term now only includes only a set of ten region fixed effects, since $ln\ c_{ni}$ is included directly. This specification mostly exploits variation of travel costs within postage cost bands to separately identify γ and η, rather than the sharp increase in postage costs across post towns on either side of a cost threshold. The correlation between travel costs and postage costs, and the fact that the first of these is likely to be measured with more error than the second, makes it more challenging to clearly separate these two elasticities.

The main message from these robustness exercises is that we find consistent evidence that the reduction in communication costs generated by the reform, which resulted in differential increases in letter market access across locations, was associated with an increase in patenting.

Table B7: Patent data analysis results: alternative θ values

	DV: Number of patents			
	Baseline: Clustered SEs.	Baseline: Bootstrapped SEs	LMA: communication with all districts	LMA: single equation estimation
Trade Elasticity θ	1	2	3	4
$\theta = 1$	0.260*	0.260	0.024	0.359*
	(0.153)	(0.180)	(0.045)	(0.209)
$\theta = 2$	0.117*	0.117	0.428***	0.445*
	(0.069)	(0.076)	(0.160)	(0.245)
$\theta = 3$	0.118*	0.118*	0.442***	0.437*
	(0.064)	(0.067)	(0.165)	(0.238)
$\theta = 4$	0.116*	0.116*	0.422***	0.378*
	(0.063)	(0.063)	(0.158)	(0.213)
$\theta = 5$	0.115*	0.115*	0.403***	0.312*
	(0.063)	(0.062)	(0.150)	(0.181)
$\theta = 6$	0.114*	0.114*	0.389***	0.271*
	(0.063)	(0.062)	(0.144)	(0.159)
$\theta = 7$	0.114*	0.114*	0.378***	0.249*
	(0.063)	(0.062)	(0.140)	(0.146)
$\theta = 8$	0.115*	0.115*	0.370***	0.237*
	(0.063)	(0.063)	(0.137)	(0.139)
$\theta = 9$	0.115*	0.115*	0.363***	0.230*
	(0.063)	(0.063)	(0.135)	(0.134)
$\theta = 10$	0.116*	0.116*	0.358***	0.225*
	(0.064)	(0.064)	(0.134)	(0.131)
$\theta = 11$	0.117*	0.117*	0.355***	0.222*
	(0.064)	(0.064)	(0.133)	(0.129)
$\theta = 12$	0.117*	0.117*	0.351***	0.220*
	(0.064)	(0.065)	(0.132)	(0.128)
$\theta = 13$	0.118*	0.118*	0.349***	0.219*
	(0.064)	(0.065)	(0.131)	(0.127)
Post Town FE	Y	Y	Y	Y
County $\times$ Decade FE	Y	Y	Y	Y
N	688	688	688	688

*See text for a detailed description of this table. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.*

C Construction of the Post Town Network

This section explains how we create the network of post roads and post towns from historical maps and records.

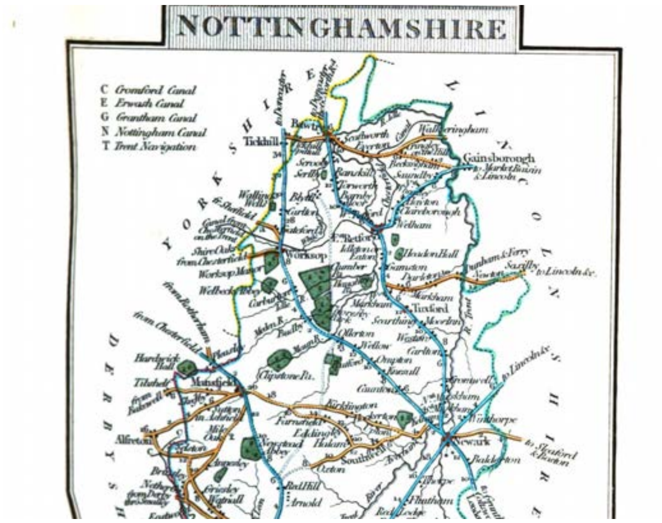
Post Towns Our list of post towns in 1838 is taken from a publication by the *Select Committee on Postage*. It lists 862 post towns and over 1,600 sub post-towns in England, Scotland or Wales that are organized into eight regional post districts.

To geolocate the post and sub post-towns, we first make use of the Google Geolocation API. Since some post town names are not unique (e.g. Bradford), this procedure comes with some imprecision. To overcome this and assess the validity of the geolocation exercise, we exploit the spatial clustering in the data that is implicitly provided by the assignment of post towns to postal districts. Specifically, we check if all coordinates from the same postal district are clustered and correct outliers manually using historical gazetteers and maps. This procedure leaves us with the centroids of all post towns and sub post-towns that are shown in Figure 3 in the main text.

Post Roads To create the post road network, we start with Cary (1828) who provides detailed maps of the road network and supplement it with the more stylized postal road network published in Basire (1838). Figure C1 shows an excerpt from both publications. Specifically, we take the locations of post towns from the previous geolocation exercise as given and connect them with contemporary B-roads that approximate the historical location of the roads shown on the maps. While this procedure introduces some measurement error from potential changes in the exact route, it comes with the important benefit that road locations implicitly take terrain into consideration. We then use post town locations and post roads to create a road network in ArcGIS that allows us to route between locations. To connect post towns to the postal route network, we create straight line minimum-distance connections between the centroid and the postal road network and assign zero-distance (and hence zero communication costs) to these connector bits. The rationale is that we consider a flat cost of one penny for sending letters within post towns, and therefore assume that every post town has direct access to the post road network.

To test how well our postal network approximates historical travel distances, we make use of information on the road distance between London (Edinburgh) and all English or Welsh (Scottish) post towns published by the *Select Committee on Postage* and compare it to calculations based on our own road network. Figure C2 shows the results of this exercise. It is reassuring to see that all observations tightly fit the 45 degree line, suggesting that our network does a good job approximating historical travel distances. The fit is particularly tight in England and a bit more noisy in Scotland. Further inspection suggest that these differences result from locations in the highlands, suggesting that in this rugged terrain, modern roads do not approximate historical roads so well. Fortunately, these differences are of little concern to

Figure C1: Historical Maps



(a) Cary 1828

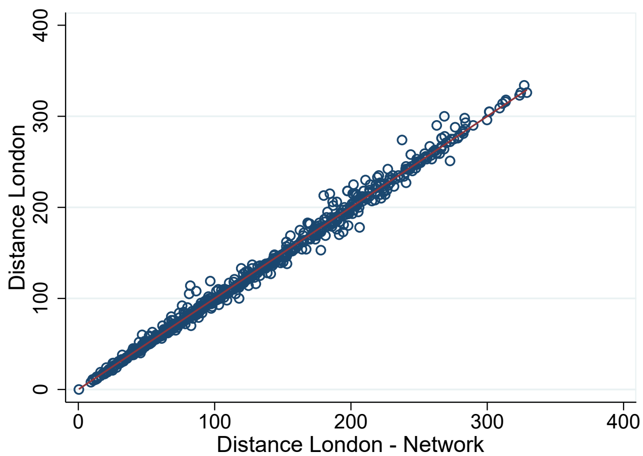
(b) Basire 1838

Note: This figure shows excerpts from Cary (1828) and Basire (1838) maps which were digitized to compute postal network routes.

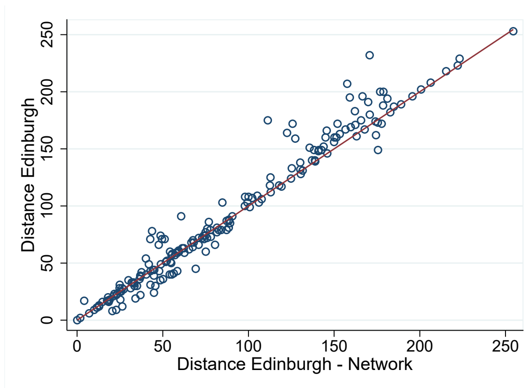
our analysis since we do not use Scotland in the main analysis of patent data, we exclude Scotland in a robustness exercise for the citation data, and we only observe very few correspondents located in the Highlands.

Connections We are ultimately interested in the costs of sending a letter between locations of scientists or inventors. To link these locations to the postal network, we assign each address location to the nearest sub post-town (15% of all cases) or post town (85% of all cases). For this purpose, we have digitized data from the Select Committee on Postage (1838, Appendix, p. 138–152) that links each sub post-town hierarchically to a post town. Put differently, it provides information about the local network of post offices. In the period before the reform, local penny posts connected these local post offices to a main post office that processed letters outside the local network. As apparent from the name local penny post, it cost one penny to send a letter within this local network. For simplicity, we assume that every letter sent was processed by a local penny post. The following Appendix Section D explains how this affects the costs of sending a letter.

Figure C2: Distance Comparisons: Historical Sources vs. Our Calculations



(a) Distance to London



(b) Distance to Edinburgh

Note: This figure shows historical distances reported in 1838 compared to the network calculations.

Table D1: Costs of sending a letter in January 1838

Mileage	Single	Double	Treble	Ounce	+ea. 1/4oz
0-15 miles	4d	8d	1s	1s 4d	+ 4d
15-20 miles	5d	10d	1s 3d	1s 8d	+ 5d
20-30 miles	6d	1s	1s 6d	2s	+ 6d
30-50 miles	7d	1s 2d	1s 9d	2s 4d	+ 7d
50-80 miles	8d	1s 4d	2s	2s 8d	+ 8d
80-120 miles	9d	1s 6d	2s 3d	3s	+ 9d
120-170 miles	10d	1s 8d	2s 6d	3s 4d	+ 10d
170-230 miles	11d	1s 10d	2s 9d	3s 8d	+ 11d
230-300 miles	1s	2s	3s	4s	+ 1s
Each extra 100 miles	+1d	+2d	+3d	+4d	+ 1d

This table shows how postage rates increase with distance. Single, double, treble refers to the number of sheets. d: penny; s: schilling; $12d = 1s$.

D Calculation of Postage Rates

This section explains how we calculated the total cost of exchanging letters between post towns. The total costs are determined by two components, a distance-dependent component that measures the costs for exchanging a letter between a pair of post towns, and a local cost component which takes delivery at the origin and destination post town into account.[5]

Distance-dependent Rates We use the distance-dependent postage rates for single-sheet letters that are provided in Postage Act (1812) and shown in Figure 1.[6] Table D1 displays the pre-reform postage rate schedule, including the rates for multiple-sheet letters.[7]

We employ our network of post towns and post routes to calculate a distance matrix that contains the distance between each possible pair of post towns. We apply the above described rates to these distances to determine the distance-dependent cost component.

The Postage Act of 1812 introduced an additional half penny for every letter that travelled through Scotland. We take this into account by adding half a penny to all connections that involve at least one Scottish post town.

Local Rates Local penny post offices were in charge of delivering the mail within post towns and

[5]Not all inventors and scientists lived directly in one of the 618 post towns.

[6]Single-sheet letters accounted for more than 90 percent of all letters in Great Britain in the pre-reform period (Select Committee on Postage, 1838).

[7]We used a fixed postage cost structure as it existed at the beginning of 1838. We abstract from some small reforms that were implemented in the period just before the main reform that we study.

connecting the surrounding smaller villages to the post towns. We add a penny at both the origin and destination location for local delivery from the local penny post office closest to the author to the origin post town and for the local delivery from the destination post town to the penny post office closest to the cited scientist. Some authors and scientists lived directly in the origin or destination post town; in these cases, we may overstate the postage. As almost all local penny posts charged a penny for single-sheet letters, we are also assuming a rate of one penny if two scientists lived in the same post town.

There are two exceptions to the above description. The London Twopenny Post and the Edinburgh Penny Post both covered a sizeable area and had specific rates that exceeded those of the other local Penny Posts. We reconstructed the elliptic coverage area of the Edinburgh Penny Post based on the description in Select Committee on Postage (1838, Appendix p. 171). The rate for using the services of the Edinburgh Penny Post was either one penny or two pence depending on the distance between the respective local penny post office and the main post office at HM Register House. We apply a rate of 1.5 pence for local delivery in and around Edinburgh.

The London Twopenny Post consisted of two separate areas with different associated rates. The town area covered all parts of London that were within a three-mile radius around the General Post Office in St Martin's Le Grand. The country area covered the suburbs outside the three-mile but within a twelve-mile radius around the General Post Office. Letters that were posted and delivered within the town area cost 2 pence. Delivery from the General Post Office to the town area or vice versa was free. Delivery from the General Post Office to the country area or vice versa was 2 pence and delivery from the town to the country area or vice versa was 3 pence as was within-country delivery (see Hemmeon, 1912).

E Citation Dataset Appendix

Our measure of scientific knowledge flows exploits information from articles published in the *Philosophical Transactions of the Royal Society of London*. The *Philosophical Transactions* was the premier British scientific journal during our study period that published articles across all branches of science. To understand how the reduction in communication costs affected communication flows between scientists, we collect the names of citing and cited scientists from this journal publications. We then use additional biographical information to identify location information for all scientists who published or were cited in the *Philosophical Transactions*. Below, we provide additional details on the data construction process, followed by some descriptive statistics of our data.

E.1 Construction of the Citation Dataset

Article submission There were two main ways that articles could be submitted to the editors of *Philosophical Transactions*. Most of the authors of our articles were affiliated with the *Royal Society of London*, which published *Philosophical Transactions*. This group was able to submit articles by communicating them directly to the editors of *Philosophical Transactions* for potential inclusion in the journal.

A subset of our authors were not not affiliated with the Royal Society. This meant that they published their articles through a slightly different process than Society members. Instead of directly communicating their findings to the editors of the *Philosophical Transactions of the Royal Society of London*, fellows of the *Royal Society of London* intermediated the correspondence between the unaffiliated authors and the editors. If a correspondent was involved in the communication during the submission, review or publication process, the header of the published article states "communicated by" followed by the correspondent's name. We identify these corresponded articles in order to identify citations between the author and the corresponding member of the Royal Society. These citations account for only a small fraction of our total citations (around three percent).

Coauthored articles In a very small number of cases, an article is published by more than one author. We observe one article published in 1844 that lists two authors and three additional articles (one published in the pre-reform period and the others published post reform) which include parts that were written by scientists other than the stated authors. We treat the one co-authored article as if each coauthor had individually written the article and, for simplicity, we count the coauthor connection as one additional citation. Accounting for co-authorship in this way increases the article count by one, the page count by 19, and leads to double-counting of eight citations.

For the three articles that contain contributions by scientists other than the stated authors, we decided to split the publication and consider those parts of the article that were not written by the stated authors as separate articles. The pages and citations were allocated accordingly. Doing so increases the article count by three, it has no effect on the page count, and results in a modest increase in citations.

Identifying scientists We begin constructing our citation dataset by collecting the names of all cited and citing scientists from the *Philosophical Transactions* articles published during our study period. Unlike modern practice, in the articles we study scientists are typically identified by their name and title, as well as some discussion of their work, but without any additional reference data. Given this, the first step in our data construction process is identifying the cited individual (e.g., obtaining their full name and other relevant information).

To identify scientists, we use the available information on a scientist's last name, area of research, and any titles or affiliations that are mentioned in a citation to seek additional biographical information that allows us to (i) uniquely identify the scientist and (ii) locate him/her at the time the article was published.

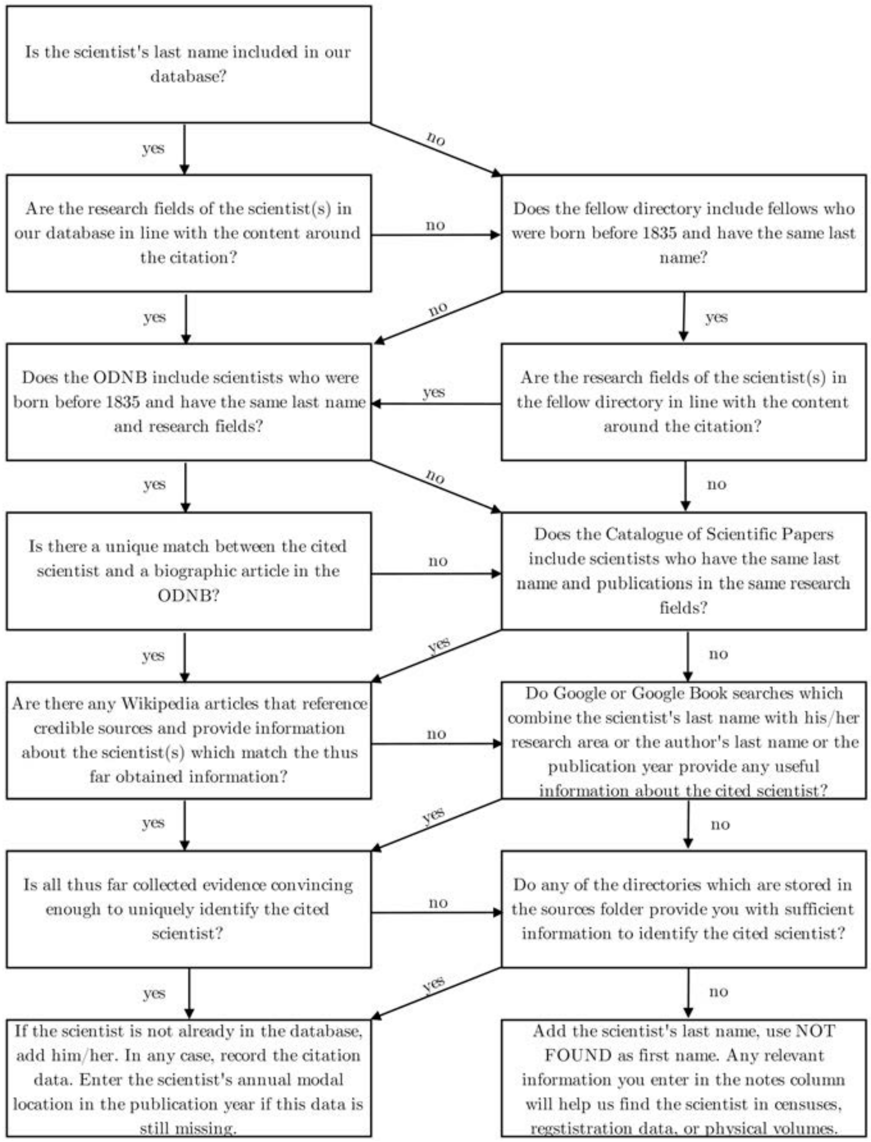
We conduct this systematic review of biographical information with the help of a team of trained research assistants who were instructed to closely follow the protocol outlined in Figure E1. In a first step, we searched the fellows directory of the Royal Society, the *Oxford Dictionary of National Biography*, and the *Catalogue of Scientific Papers* for biographic information. If none of these resources allowed us to uniquely identify the full name of the scientist, we extended our efforts to Google searches for other web-based resources which could then be compared to the available information on the scientist's work, as described in the citing article.

Locating Scientists Once we have uniquely identified all scientists, we collect additional biographical information including the scientist's institution and a geolocated work address. We start our search for location information in the same sources that helped us to identify the scientists and extend the search to additional sources whenever necessary. For each publication year, we were able to identify and geolocate at least 97.2 percent of all cited scientists, averaging 98.9 percent in total. This high success rate is mainly due to three reasons: the influence and popularity of the authors and cited scientists, the plethora of historical information on these scientists that is readily available online, our systematic approach for navigating this information, and the considerable timely effort we devoted to this data collection.

Our geocoding step uses several trade directories, the 1841 Census, scientists' preserved individual correspondence, and lists covering specific professions. The latter include Clifton's (1995) directory of British scientific instrument makers, O'Byrne's (1849) dictionary of officers in the Royal Navy, and the British Almanacs (1828-1875) which list all university professors in the UK.

Figure E2 shows three text snippets from a biographic article on Richard Owen (1804-1892) to illustrate the type of information we are looking for. For this specific example, we would note as institution *Hunterian Museum, Royal College of Surgeons* and Google provides the address as *38–43 Lincoln's Inn*

Figure E1: Data Digitization Procedure



Note: This flow chart summarizes the process to uniquely identify the scientists who are cited in the Philosophical Transactions.

Fields, London WC2A 3PE. This address information is easy to geolocate. If the institution no longer exists, we search historical sources for address information. In cases where an institution was located in a small village or a specific part of town with no exact address, we use the centroid of the location as address coordinates. In some cases, the address of an institution such as the Huterian Museum may have changed over time. While this is unlikely to be an issue because our analysis is at the town level and it would be very rare for institutions to change city, we have carefully checked to make sure that the institutions used in our location data have not done so.

Figure E2: Extracts of text from original biographical source

In the mid-1820s the Royal College of Surgeons was in a difficult position. It found itself one of the

Owen's entry into the Hunterian Museum on 7 March 1827 at a salary of £30 per quarter was the

Although in 1832 Owen's career remained uncertain, the five years as junior curatorial assistant at the Hunterian provided the foundation for his future career. The particular investigations he initiated and

Note: This figure shows three extracts from a biographical article about Richard Owen (1804-1892), illustrating the information we look for in geolocating scientists.

In rare occasions, scientists were travelling in the year when we observe them. If the scientist's modal location for a given year is his/her regular address in Great Britain, we adopt it. Otherwise, we drop the observation because our focus lies on communication flows within Great Britain.

E.2 Summary Statistics

Between 1830 and 1849, a total of 443 articles were published in the *Philosophical Transactions of the Royal Society of London*. Among these articles, 97 percent had at least one citation and only 15 articles do not contain any citations. Table E1 displays basic descriptive statistics for this population at the year, period, and aggregate level.

We observe that about 57 percent of all articles and pages were published before the introduction of uniform penny postage. Articles continued to span about twenty pages on average but journal space became more evenly distributed among authors over time.

In our study period, it was not uncommon for an individual author to publish multiple articles in the same issue. This explains why the number of articles exceeds the number of authors in most years. As mentioned above, co-authorship was rare, but some articles still represent joint work. These articles typically state a single author but include notes by other scientists that often extend to several pages. Table E1 only takes official co-authorship into account which occurred in one article published in 1844. In addition, there are three articles in the sample where sizable parts were written by scientists other than the stated authors. As mentioned above, we split these articles up and attribute each forward citation to

Table E1:
Articles in the *Philosophical Transactions* in the observation period

Period	Articles	Pages	Article length	Authors	Articles per author	Pages per author
1830	26	419	16.12	20	1.30	20.95
1831	28	495	17.68	18	1.56	27.50
1832	26	599	23.04	17	1.53	35.24
1833	29	811	27.97	24	1.21	33.79
1834	29	582	20.07	23	1.26	25.30
1835	21	356	16.95	18	1.17	19.78
1836	26	605	23.27	20	1.30	30.25
1837	24	430	17.92	19	1.26	22.63
1838	20	399	19.95	13	1.54	30.69
1839	22	420	19.09	16	1.38	26.25
Pre period	251	5116	20.38	99	2.54	51.68
1840	24	604	25.17	20	1.20	30.20
1841	21	297	14.14	16	1.31	18.56
1842	15	300	20.00	15	1.00	20.00
1843	14	327	23.36	13	1.08	25.15
1844	12	319	26.58	13	0.92	24.54
1845	17	357	21.00	13	1.31	27.46
1846	29	626	21.59	22	1.32	28.45
1847	17	251	14.76	12	1.42	20.92
1848	19	269	14.16	17	1.12	15.82
1849	24	507	21.13	20	1.20	25.35
Post period	192	3857	20.09	93	2.06	41.47
Total	443	8973	20.26	161	2.75	55.73

the split origin. Doing so increases the article counts by one to a total of 252 in the pre- and by three to a total of 195 in the post-reform period. This increases the total count to 447 articles. The split of the two coauthored articles which fall into the post-reform period increases the number of distinct authors in that period to 95 and the total number of distinct authors in all periods to 163.

We are interested in citation pairs that connect scientists who lived in Great Britain in the publication year of the respective article. This means we can restrict our search to articles that were written by authors who resided in Great Britain in the publication year(s) of their article(s). Table E2 zooms in on these articles. The coauthored articles are already split up in Table E2. Figure E3 relates the annual article count produced by authors who resided in Great Britain to the total annual article count. It shows that most of our articles were authored by scientists who resided in Great Britain.

Figure E3: Annual Article Counts

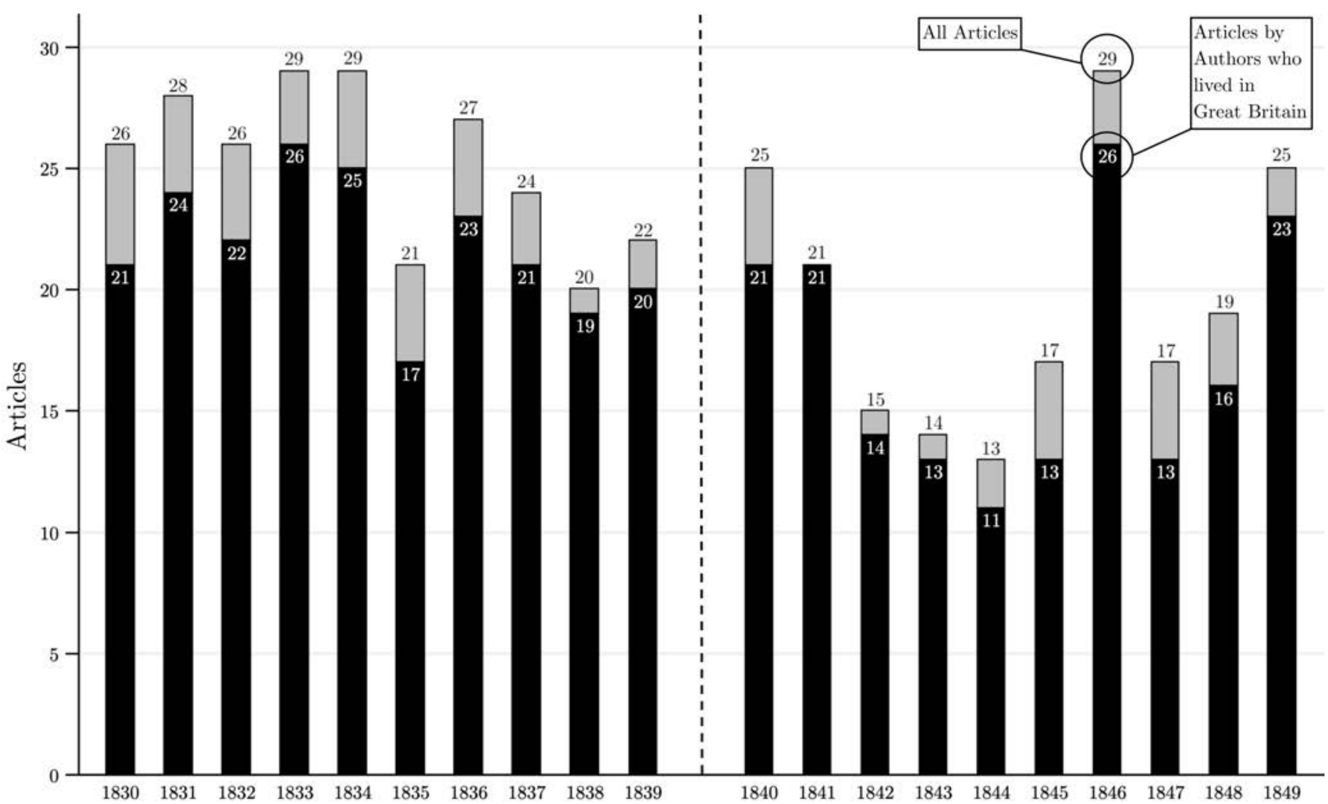
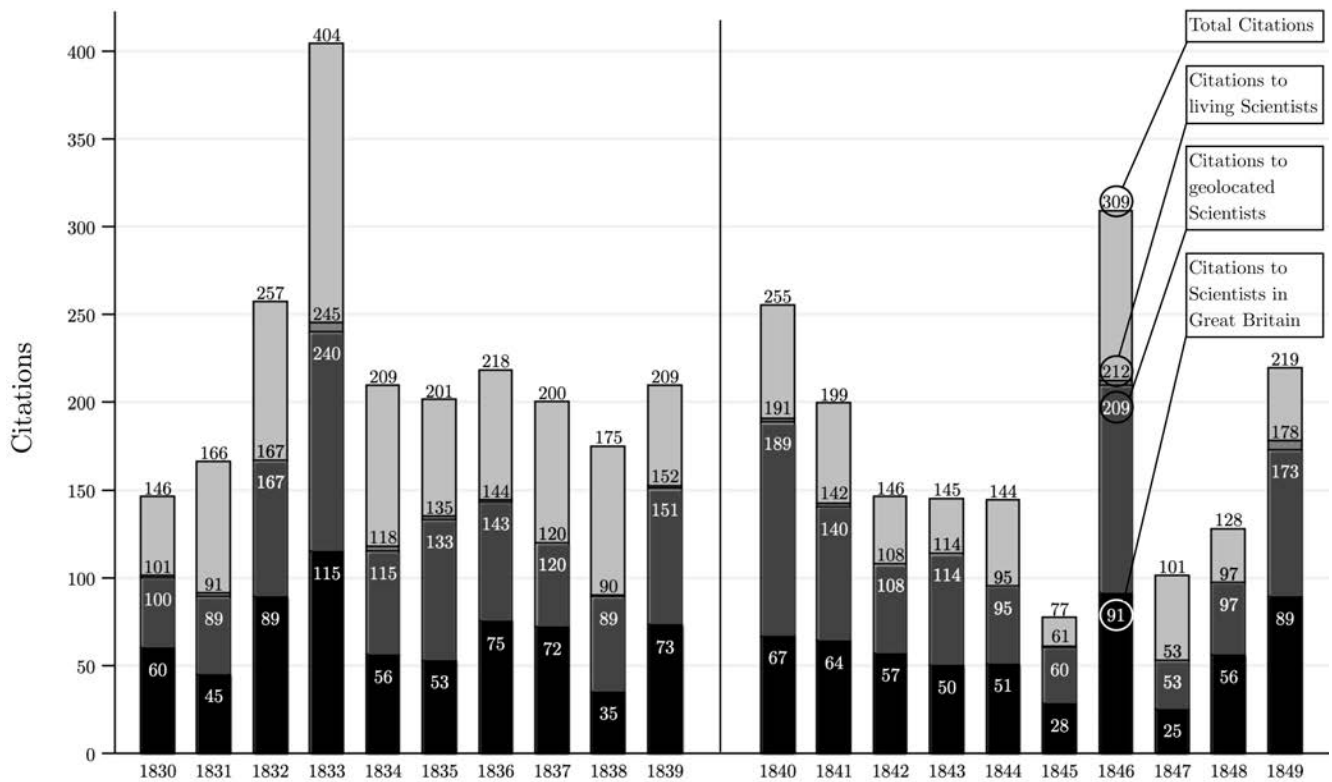


Figure E4 displays summary statistics for all citations in these articles. The black bars represent the

Table E2:
Articles by Authors who lived in Great Britain in the Publication Year

Period	Articles	Article share	Pages	Page share	Article length	Authors	Author share	Articles per Author	Pages per Author
1830	21	80.8	358	85.4	17.0	17	85.0	1.24	21.1
1831	24	85.7	408	82.4	17.0	16	88.9	1.50	25.5
1832	22	84.6	560	93.5	25.5	17	100.0	1.29	32.9
1833	26	89.7	791	97.5	30.4	24	100.0	1.08	33.0
1834	25	86.2	490	84.2	19.6	22	95.7	1.14	22.3
1835	17	81.0	288	80.9	16.9	15	83.3	1.13	19.2
1836	23	85.2	533	86.9	23.2	20	100.0	1.15	26.7
1837	21	87.5	405	94.2	19.3	18	94.7	1.17	22.5
1838	19	95.0	358	89.7	18.8	13	100.0	1.46	27.5
1839	20	90.9	394	93.8	19.7	16	100.0	1.25	24.6
Pre period	218	86.5	4585	89.5	21.0	89	89.9	2.45	51.5
1840	21	84.0	521	86.3	24.8	19	90.5	1.11	27.4
1841	21	100.0	297	100.0	14.1	16	100.0	1.31	18.6
1842	14	93.3	266	88.7	19.0	14	93.3	1.00	19.0
1843	13	92.9	291	89.0	22.4	13	100.0	1.00	22.4
1844	11	84.6	317	93.8	28.8	11	84.6	1.00	28.8
1845	13	76.5	309	86.6	23.8	13	100.0	1.00	23.8
1846	26	89.7	594	94.9	22.8	22	100.0	1.18	27.0
1847	13	76.5	222	88.4	17.1	12	100.0	1.08	18.5
1848	16	84.2	224	83.3	14.0	16	94.1	1.00	14.0
1849	23	92.0	470	92.7	20.4	20	95.2	1.15	23.5
Post period	171	87.7	3511	90.6	20.5	88	92.6	1.94	39.9
Total	389	87.0	8096	90.0	20.8	150	92.0	2.59	54.0

Figure E4: Annual Citations by Authors who lived in Great Britain in the Publication Year



set that will feature in our analysis: those with living scientists who were resident in Great Britain at the time of publication. When these citations are from an article by an author who also lived in Great Britain, the combination provides a bilateral connection between scientists within the country. The set of these bilateral connections forms our analysis dataset.

Table E3 displays the number of citations collected for all years and how different sets of restrictions reduce our sample. The first column shows the number of *Total Citations* we have entered for each year. This column excludes 73 references in an 1836 article to military and other personnel who conducted measurements and are only listed in the tables that display these measurements (see Whewell, 1836). We exclude these individuals because, unlike regular scientists, they are named in the *Philosophical Transactions* only because they were assigned to a particular station where they were responsible for data collection.

The next column, *Citations to living scientists*, reduces the sample to scientists who were alive in the publication year. The column *Citations to geolocated scientists* reduces the sample to cases where we were able to find the locations of the author and the cited scientist. Finally, the column *Citations to scientists in Great Britain* shifts our focus to citations that connect two scientists who resided in England, Wales, or Scotland in the publication year of the article which contains the citation. This pair of columns reflects the set of citations to scientists in Great Britain by authors also located in Great Britain, the set of citation connections that form our analysis data set. We can see that about half of all of the citations by British-based authors were to other scientists located in Great Britain.

The respective base of each percentage share is the category to its left. The share of citations to living scientists is given as a percentage share of all citations; the share of citations to geolocated scientists is given as a percentage share of the citations to living scientists; and the share of all citations to scientists in Great Britain is given as a percentage share of the citations to geolocated scientists. The high percentage of geolocated scientists once again reflects the considerable timely effort that we devoted to the data collection.

Table E3:
Citations by Authors who lived in Great Britain in the Publication Year

Period	**Total citations**	**To living scientists**		**To geolocated scientists**		**To scientists in Great Britain**	
	count	count	share	count	share	count	share
1830	146	101	69.2	100	99.0	60	60.0
1831	166	91	54.8	89	97.8	45	50.6
1832	257	167	65.0	167	100.0	89	53.3
1833	404	246	60.9	241	98.0	115	47.7
1834	209	118	56.5	115	97.5	56	48.7
1835	201	135	67.2	133	98.5	53	39.8
1836	218	144	66.1	143	99.3	75	52.4
1837	200	120	60.0	120	100.0	72	60.0
1838	175	90	51.4	89	98.9	35	39.3
1839	209	152	72.7	151	99.3	73	48.3
Pre period	2,185	1,364	62.4	1,348	98.8	673	49.9
1840	255	192	75.3	190	99.0	67	35.3
1841	199	142	71.4	140	98.6	64	45.7
1842	146	108	74.0	108	100.0	57	52.8
1843	145	114	78.6	114	100.0	50	43.9
1844	144	95	66.0	95	100.0	51	53.7
1845	77	61	79.2	60	98.4	28	46.7
1846	309	212	68.6	209	98.6	91	43.5
1847	101	53	52.5	53	100.0	25	47.2
1848	128	97	75.8	97	100.0	56	57.7
1849	219	178	81.3	173	97.2	89	51.4
Post period	1,723	1,252	72.7	1,239	99.0	578	46.7
Total	3,908	2,616	66.9	2,587	98.9	1,251	48.4

F Construction of the Patent Dataset

Our patent data were compiled by the British Patent Office (Woodcroft, 1854) as part of the 'Titles of Patents of Invention' and subsequently digitized Nuvolari & Tartari (2011). We observe a total of 7,905 patents over our study period from 1830–1849 with the number of patents ranging between a minimum of 168 in 1831 and 608 in 1845. 260 of these patent applications show locations outside Great Britain and 29 applications have a missing location. This leaves us with a sample of 7,616 observations in Great Britain to be matched with post towns or sub post-towns.

To assign patents to post towns or sub post-towns, we proceed in multiple steps. First, we manually clean all address information. For instance, a large number of addresses show up as London street names without explicit mention of London, so we assign London as location. We then move on and match the location information in the patent data with the location register of post town and sub post-town locations. In this step, we match 86% of all Great British patents (6,570 our of 7,616) in our sample period to a post town or sub post-town.

To assign the remaining 14% of patents to post towns or sub post-towns, we first geolocate all remaining 1,046 locations using the Google geocoding API. We carefully checked the results of the geolocation process and conducted manual location searches if google did not provide satisfactory results. This procedure provides us with coordinates for all remaining locations where we observe patent applicants. We use this location information to calculate an applicant's distance to all post towns or sub post-towns and we assign them to the nearest (sub) post town. The average distance to the next (sub) post town is 4km; at the median, locations are within 3.4km of the next (sub) post town; and at the 90th percentile, they are 8.4km away from the next (sub) post town. Locations that are further away are not relevant since they are located in the Scottish Highlands and we do not consider Scotland in our patent analysis because of the differences in the patent system described in the main text.[8] At the end of the matching procedure and after dropping Scotland, we end up with 7,172 patent applications from locations in England and Wales.

[8]The reason why we do not drop locations in Scotland before is that the patent data do not come with a country identifier. As a result, we only identify locations in Scotland once we have matched them to post town or sub post-town in Scotland.