A Penny for your Thoughts

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

November 11, 2021

Preliminary draft; 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 new ideas and technologies are developed.

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

*Northwestern University, Department of Economics
†University of Toronto, Munk School of Global Affairs and Public Policy and Department of Economics
‡Georgetown University, McDonough School of Business
§Vanderbilt University, Department of Economics

1 Introduction

The exchange of knowledge is thought to play a central role in the development of new ideas and economic growth (Romer, 1986; Lucas, 1988). In fact, communication costs are considered so important that they are often cited as a primary explanation for the existence of cities (Marshall, 1890; Duranton & Puga, 2004; Davis & Dingel, 2019). Existing work using patent citations suggests that knowledge flows decline with distance, with implications for the development of new ideas and technologies (see, e.g., Jaffe *et al.* (1993)). However, isolating the causal impact of communication costs on innovation rates has proven challenging. This is particularly true in modern environments, where knowledge can be exchanged through a wide variety of channels.

One promising avenue for overcoming this challenge is to focus on a historical environment with fewer communication channels. In this paper, we leverage a historical setting characterized by a very large and spatially-varied change in the cost of the primary form of long-distance communication in order to examine the causal relationship between communication costs and the development of both scientific and technological knowledge, two outcomes that are of particular interest because of the role they play in economic growth.[1] Specifically, we study the "Uniform Penny Post" introduced in Britain in 1840 (Postage Act, 1839). 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. Because the new uniform charge replaced a distance-based postage system, this reduction was also spatially varied. We leverage these features in order to estimate the impact of a change in communication costs on the development of new ideas and new technologies.

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 between 88 and 100 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.

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 operating in England, Wales or Scotland just before the reform, over 1600 sub-post offices, and the postal road network through which each post

[1] This approach has been taken in a number of recent papers examining how communication costs affect outcomes such as trade flows or financial markets (Koudijs, 2014; Steinwender, 2018; Juhasz & Steinwender, 2018). However, existing studies have not examined the impact on the development of new scientific knowledge or technological development.

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 between two-thirds and three-quarters of the decay of citations with distance across locations. This provides a first piece of evidence on the impact of communication costs on scientific knowledge flows.

To complement our evidence on knowledge flows between scientists, in the second part of the analysis 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 is of particular importance for economic growth over the short to medium term. Our analysis takes advantage of geolocated patent data covering thousands of patents filed between 1830 and 1850. There is one important difference between our patent data analysis approach and what we do with the citation data. In particular, 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. 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, so our patent data analysis will always include a standard market access measure, calculated over the turnpike road network, as a control.

To implement this research design, we need to estimate the increase in towns' letter market access caused by the reform, for which elasticities of letter flows to distance, cost and population are needed. We recover these elasticities using data on letters sent to London from hundreds of post towns in one week in 1838. Using a simple difference-in-difference analysis 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 the proximity to railway stations. 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 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 on other economic outcomes, including Jensen (2007), Goyal (2010), Aker (2010), Allen (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 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.[2]

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.* (2016) 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. 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.[3] This makes it hard to isolate the role of communication from other omitted local factors, as well as other impacts of proximity such as reduced transport costs.

[2]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.

[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).

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.

Finally, our paper improves our understanding of a key event in British economic history. As the economist John Ramsey McCulloch argued in 1833, "nothing contributes more to facilitate commerce than the safe, speedy and cheap conveyance of letters."[5] Our evidence suggests that, in addition to facilitating commerce, the introduction of the uniform penny post also contributed to the exchange and development of technological and scientific knowledge. This provides a particularly interesting example of how this institutional reform contributed to sustaining technological progress and economic growth during the Industrial Revolution. 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.[6] Moreover, 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).

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. This is followed by Section 4 that establishes the relationship between of 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,

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

[5]Quoted from Robinson (1948).

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

comprise over 4,900 letters (James, 1991-2011). So important and voluminous were these correspondence 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).

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.[7] 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.

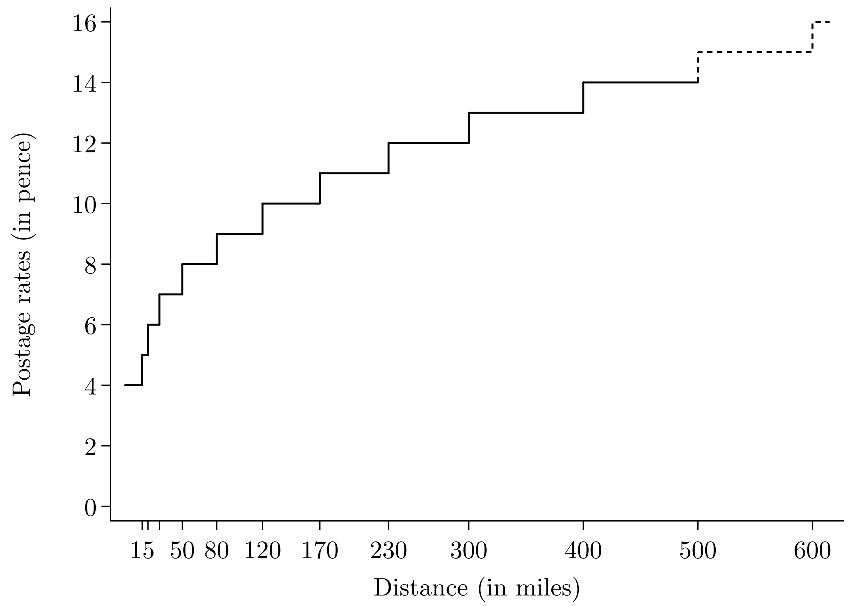
In the late 1830s, Rowland Hill became a leading advocate of post office reform. In 1837, he published his famous pamphlet, *Post Office Reform, its Importance and Practicability,* in which he argued for the introduction of a flat postage rate independent of distance for all letters up to 14 grams, and the introduction of prepaid postage indicated by a stamp on the letter.[8] 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

[7]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. Non-commissioned officers in the Royal Navy and Royal Army already benefited from this low uniform rate before the reform. In our analysis below, the presence of these privileges will reduce the impact of treatment, since anyone with franking privileges and any non-commissioned officer in the military will not be treated by our reform. This will have the effect of pushing our estimated treatment effects toward zero, though we do not expect the effect to be very large. Thus, our results can be thought of as a lower bound on the impact of the reform had it applied equally to everyone in the country.

[8]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.

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.

1840.[9] 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).[10] 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%.[11]

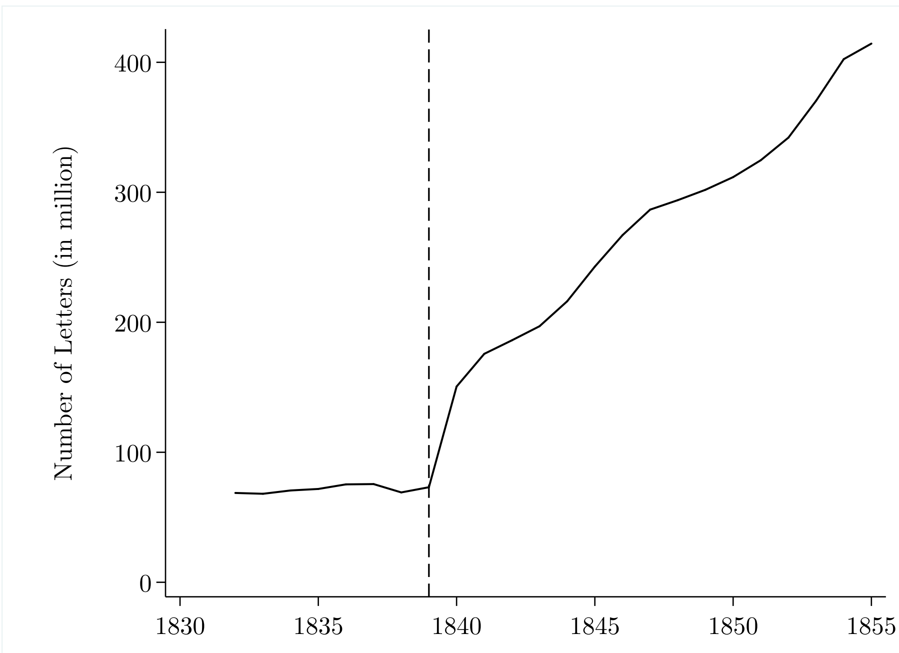
[9]Some MPs were initially skeptical that Hill's plan would generate sufficient revenue and called for a higher uniform rate. The resulting Uniform Fourpenny Post launched on 5 December 1839, led to a volume increase of 28 percent within less than a month, convinced skeptics of the uniform penny rate,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

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

[11]The calculation of the lower bound relies on Colquhoun (1806). For the upper bound, a salary of £300 per year is used

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. [12] Since few alternatives existed for long-distance communication, such increase in the volume of letters likely represents a net increase in communication flows within the country. In response to the rise in letters, the postal system expanded rapidly, from 4,028 post offices in 1840 to 11,235 in 1858 (General Post Office, 1855, p.20; General Post Office, 1859, p.7). Greater letter volume also facilitated an increase in the frequency of deliveries, which improved the convenience of using the post.[13] 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



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. John Henslow, a

based on a figure listed in the Oxford Dictionary of National Biography for a chaired professor at the University of Cambridge (see Chapman, 2011).

[12] In Great Britain, 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 1839 and 1849 by the 1841 and 1851 census population data.

[13] See Select Committee on Postage (1843) pp.258-261.

friend and mentor of Charles Darwin, wrote that "To the importance of the penny postage to those who cultivate science, I can bear most unequivocal testimony, as I am continually receiving and transmitting a variety of specimens by post. Among them, you will laugh to hear that I have received three living carnivorous slugs, which arrived safely in a pill-box!" (Lewins, 1865, p.200). This quote nicely summarizes why we would expect this sudden decrease in communication costs to have a marked effect on science and technological progress.[14]

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 railway system in our analysis. The telegraph was also introduced, starting in 1844, but initially this was used just as a signaling system for the railways so it is less of a concern for our study.[15]

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 network (which is not the same as the turnpike network, since not all roads were post roads) and connected them to a digitized and geocoded list of 618 post offices.[16] Appendix C contains the details of this step. Figure 3 presents our digitized map of postal roads and post offices.

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.

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

[14] 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.

[15] 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.

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

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, Wales and Scotland separated by white border lines. Panel (a) shows post towns (black dots) and sub-post towns (gray cross) observed around 1838. Panel (b) shows post towns (black dots) and postal roads digitized from maps around 1838.

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 distance-based, but that this took the form of a step function. 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.[17] 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.

Our analysis requires citation, location, and biographical data for all scientists who published or were cited in the *Philosophical Transactions* in the observation period.[18] 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. [19] An example of a citation from a 1840 article in the *Philosophical Transactions* is reported in figure A1 in the Appendix.

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.[20] We then follow a systematic approach to identify and geolocate all scientists and inventors who are cited in this set of 389

[17] Letters with more pages faced higher rates but a similar step function increase in the cost at the same distance bands.

[18] A detailed description of the lengthy and labor-intensive data collection process is included in the Appendix D.

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

[20] Appendix D.1 provides additional information on both sets of articles. Summary statistics are available in tables B10 and B11.

articles. We record and geocode the modal locations of these cited individuals in the publication years of the corresponding articles.[21]

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.[22] 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.[23]

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 2,540 citations, 1,219 of which referred to a scientist living in England, Wales or Scotland, with at least 88 percent of citations geolocated in every year. Table B12 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 these 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 apply.

Of course, direct correspondence through the post is only one way that one scientist may have learned about the work of another. Other channels include reading about their work in scientific publications, including the *Philosophical Transactions*, or correspondents 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. If cheaper postage facilitated this communication then we may observe more articles with citations across longer distances in the post-reform period not only because it was easier for individuals to learn about the work of other scientists who were further away, but also because more correspondence between them may have aided in the development of new ideas leading to a publication in the *Philosophical Transactions*.[24]

[21]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.

[22]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. The algorithm is included in the data appendix.

[23]The location data entry instructions that we provided to our group of research assistants are included in the data appendix. The resulting data were carefully checked for accuracy.

[24]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.[25]

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. For our purposes, it is useful to note that filing a patent was a very expensive process: Sullivan (1989) calculates that patent fees in 1830 were at least four times the average annual income. Although the postal reform might have reduced the cost of filing a patent via mail, the incidence of these savings on the total filing costs is likely of little significance.

For consistency with our citation analysis, our patent analysis focuses on the period from 1830-1850. 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.[26] 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.[27] We then link each patent to the nearest post town, either directly or through a closer sub-post town. The result is a list of 7,677 patents filed between 1830 and 1850, each assigned to a post town (our unit of analysis).[28] Appendix Figure A2 describes the number of patents in each year across the study period.

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

[26] 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).

[27] 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).

[28] 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. We have identified patent agents by reviewing personal biographies for

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, distance, postage cost, and population. 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 location treatment in the patents analysis relies on estimates of the relation between letters, population, distance and cost that will be recovered in this exercise.

Let L_n be the total number of letters originating from location n. We posit that the flow of these letters sent to a location i follows 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 turnpike distance 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 two places are farther apart; 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 distance and costs when $\gamma,\ \eta > 0$: hence, a larger LMA indicates that location n's residents have a greater opportunity to communicate at cheap prices with places which tend to be more attractive communication destinations.

Taking logs, we obtain the following estimation equation

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

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 origin and destination fixed effects. Unfortunately, bilateral letter flows were never collected in Great Britain around the time of the reform. However, we have discovered a set of data on letter flows into London from 622 origin locations for one week in January 1838 (the pre-reform period).[29] To use this

those individuals with numerous patents. A number of these turn out to be patent agents. We exclude patents with patent agents listed as the inventor from our analysis.

[29] We focus specifically on regular letters, omitting from our analysis "privileged" letters (e.g., those sent for free by those with franking privileges) as well as newspaper flows, which were governed by a separate set of policies, though we do use

information, we have to fix $i = London$ and the letter-destination fixed effect will enter the regression as a constant. In addition, we cannot use letter-origin fixed effects, as we would have as many fixed effects as observations. On the other hand, we have a measure of the turnpike distance d_{ni} between location n and London, and we can accurately measure the cost of sending letters to London, c_{ni}.

To make progress, we assume that the propensity of a place to receive or send letters is a power function of local population, that is, $L_n = \beta_0 P_n^\beta$ and $A_i = \beta_1 L_i = \beta_1 \beta_0 P_n^\beta$ In the empirical specifications below, we control for origin-location characteristics with a set of ten regions fixed effects, different approximations for the LMA term, and controls.

We adopt two alternative approaches to estimating our coefficients of interest: a simple cross-sectional regression, and a two-steps procedure that exploits the sharp increases in prices around the distance cutoffs.

In the first approach, we estimate the coefficients on population, distance and cost jointly. In particular, we estimate

$$\ln L_{ni} = \alpha_0 - \gamma \ln d_{ni} - \eta \ln c_{ni} + \beta \ln P_n + \alpha' X_n + \varepsilon_n \tag{4}$$

In this expression, d_{ni} is the turnpike distance between post towns n and i, c_{ni} is the cost of sending letter pre-reform, and P_n is the population of the registration district in which the post town is located.[30] The term ε_n captures the sum of classical measurement error in the regressors. The controls X_n include a proxy for LMA

$$\ln \widehat{LMA}_1 = \ln \sum_{i' \neq n} P_{i'} d_{ni'}^{-1} c_{ni'}^{-1} \tag{5}$$

a set of region fixed-effects, and the average distance of location n to the two closest rail stations. The proxy $\widehat{LMA}_1$ discounts other post towns' population based on the turnpike distance and the cost of sending letters. This measure is closer to equation 2, which incorporates the fact that in our data not all registration districts have post towns, and hence no bilateral communication is possible with residents of those districts. In the Appendix, we reproduce our results with an alternative proxy for LMA constructed as

$$\ln \widehat{LMA}_2 = \ln \sum_{i' \neq n} P_{i'} d_{ni'}^{-1} \tag{6}$$

where we use the entire set of registration districts in Great Britain without discounting by communication costs. This formulation departs more from equation 2, but has the potential to better capture features of the local economy that would be absorbed by an origin fixed-effect.[31] The average distance to the two

those as controls in some of our analysis. Summary statistics for our letter flow data, and other variables used in our analysis of letter flows, are available in the Appendix, Table B1.

[30] 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.

[31] We have also experimented with a hybrid of these two possibilities, a restriction of $\widehat{LMA}_2$ only to registration districts with post town, finding similar or stronger results in the patents analysis below.

closest rail stations proxies for alternative individual mobility opportunities. In the Appendix, we also show our results using the distance to the closest station, or to the closest three stations.

Our estimates of equation 4 are presented in Table 1. Column 1 looks at the relationship between distance and letter flows controlling only for region fixed effects, postage costs and turnpike distance. In Column 2 we add in controls for a location's population, then the proxy for access in Column 3, and the average distance to the two closest rail stations in Column 4. As each additional control is added, the separate impact of distance and costs on letter volumes comes into focus. As expected, locations with a larger population send more letters to London. Locations 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 and distance to London). In our preferred specification, Column 4, we estimate a distance elasticity of -0.63, which suggests that, conditional on cost, a doubling of the distance is associated with around a 47% decrease in the volume of letter flows. Conditional on distance, a doubling of the costs is associated with a 92% reduction of the volume of letters. Perhaps not surprisingly, distance and cost have a substantial and distinct impact on letter flows. Table B2 in the Appendix reproduces the most stringent specification of Column 4 using the alternative proxy for LMA and proximity to a varying number of stations.

This above cross sectional approach imposes a constant elasticity of distance and cost across all the range of distances in our sample. Further, it does not fully leverage the sharp discontinuity implied by the step-wise nature of the cost of sending letters pre-reform. In a second exercise, we adopt a two-steps procedure. In a first step, we identify the effect of distance and population while flexibly controlling for costs. In particular, we estimate

$$\ln L_{ni} = \alpha_0 - \gamma \ln d_{ni} + \beta \ln P_n + \alpha' X_n + \varepsilon_n \tag{7}$$

where X_n include the proxy for letter access $\widehat{LMA}_1$, a set of region fixed effects and cost brackets fixed effects, and the average distance to the closest two rail stations. Since the cost for sending letters is constant within brackets, any remaining relationship between distance and letter flows is independent of the cost of sending letters. Table 2 reports the results. As above, the effect of distance on letters to London become clear after we control for alternative destination options as captured by the proxy for LMA. Using cost bracket fixed-effects rather than the log cost in this first step leads to a larger distance elasticity of -0.86, and a slightly larger role for population. In Table B3 in the Appendix, we reproduce the last Column of this table for the different proxy for LMA and proximity to a varying number of stations.

In the second step, 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 changed.[32] It is worth noting that our data include towns in every direction from

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

Table 1: Impact of cost, distance, and population on letter flows

	DV: Log Letter Flows to London			
	1	2	3	4
Ln Postage Cost	-1.980	-1.667	-2.460**	-2.487**
	(1.229)	(1.088)	(1.069)	(1.077)
Ln Turnpike Distance	0.098	-0.153	-0.623*	-0.631*
	(0.332)	(0.302)	(0.322)	(0.322)
Ln Population		1.088***	1.195***	1.188***
		(0.093)	(0.098)	(0.097)
Ln $\widehat{LMA}_1$			-1.015***	-1.096***
			(0.232)	(0.247)
Ln Rail Distance				-0.052
				(0.060)
R^2	0.13	0.34	0.36	0.36
N	568	568	568	568
Regions FE	Y	Y	Y	Y

Ln Rail Distance is the log of the average distance to the two rail stations closest to the post town. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

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:

$$\ln L_{ni} = \beta_0 - \mathbf{1}[n \in G_B] + \eta \ln c_{ni} + \beta' X_n + \varepsilon_n \tag{8}$$

where $\mathbf{1}[n \in G_B]$ is a set of group dummies, and X_n include $n-$specific controls like population, the proxy for letter access $\widehat{LMA}_1$, average distance to the two closest stations, and the number of newspapers and privileged letters or packets sent to London. The group bins flexibly control for different distances to London across sets of towns. Within each group, locations are 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 set of results offers a less parametric approach to controlling for distance. However, it also relies on a substantially smaller set of observations. The results of this second step are reported in Table 3.

We estimate cost elasticities ranging between -1.2 and 1.6; these estimates become more precise as

next.

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

	DV: Log Letter Flows to London			
	1	2	3	4
Ln Turnpike Distance	-0.020	-0.303	-0.783**	-0.856**
	(0.360)	(0.331)	(0.352)	(0.353)
Ln Population		1.138***	1.216***	1.200***
		(0.094)	(0.096)	(0.094)
$Ln\ \widehat{LMA}_1$			-0.994***	-1.222***
			(0.252)	(0.285)
Ln Rail Distance				-0.122*
				(0.071)
R^2	0.14	0.36	0.38	0.38
N	568	568	568	568
Letter cost FE	Y	Y	Y	Y
Region FE	Y	Y	Y	Y

Ln Rail Distance is the log of the average distance to the two rail stations closest to the post town. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

controls are added. Column 1 only includes group threshold and region fixed-effects. Column 2 controls for local population. In Columns 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 distance, 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 distance, 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 other local factors other than cost that influenced letter flows. Columns 4 and 5 control for our proxy for LMA and proximity to railroad stations. In our preferred estimate, the elasticity of letter cost to volume of letters is estimated at about -1.6. Table B4 in the Appendix reproduces the most stringent specification of Column 5 using the alternative proxy for LMA and proximity to a varying number of stations.

Table 3: Impact of cost on letter flows

	DV: Log Letter Flows to London				
	1	2	3	4	5
Ln Postage Cost	-1.230	-1.789	-1.390	-1.516*	-1.572*
	(1.719)	(1.537)	(0.868)	(0.878)	(0.872)
R^2	0.29	0.42	0.81	0.82	0.82
N	98	98	98	98	98
Ln Population		Y	Y	Y	Y
Ln Privileged Letters			Y	Y	Y
Ln Newspaper Flows			Y	Y	Y
$Ln\ \widehat{LMA}_1$				Y	Y
Ln Rail Distance					Y
Group threshold FE	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y

Ln Rail Distance is the log of the average distance to the two rail stations closest to the post town. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

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 independent of distance, which also independently reduces letter flows. Second, this analysis provides specific estimates of the elasticity of letter flows with respect to postage costs, distance, and population. 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 transport costs affect trade flows, an analog to our interest in how communication costs affect knowledge flows.

Using n subscripts to denote origin locations and i subscripts to denote destinations, our primary regression specification is,

$$CITE_{nit} = \beta COST_{ni} * POST_t + \xi_{nt} + \gamma_{it} + \psi_{ni} + \epsilon_{nit} \quad (9)$$

where $CITE_{nit}$ 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 (and similarly, the publication of an article in a destination may lead to citations by subsequent articles in a number of origin 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 service provision of local penny posts, the presence of individuals with franking privileges at an origin or destination location, and potential effects of international rate changes. 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.

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.[33] We then estimate our specification using PPML. Note that the dyad fixed effects in Eq. 9 will cause origin-destination pairs with equal citation counts in both the pre- and the post-reform periods to drop out of the analysis.[34]

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 x 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 roughly two-thirds to three-quarters of the penalty imposed by higher costs in the pre-reform period (with the difference likely due to other impacts

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

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

of distance on citations that were independent of cost). Put another way, our results indicate that roughly two-thirds to three-quarters of the decay of knowledge exchange with distance, 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, the inclusion of this large set of additional fixed effects has very little impact on the estimated effect of the postal reform. 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 this very stringent specification results in very little change in the estimated effect of the reform on bilateral citations.

In Columns 4-6 we estimate similar results but using an alternative approach to calculating standard errors. Specifically, in place of the robust standard errors used in Columns 1-3, we cluster standard errors at the origin-period level.[35] This allows for the possibility that standard errors may be correlated across dyads within a period because the citations originated from the same article. Allowing for this type of correlation has very little impact on the statistical significance of our results.

The estimated coefficients in Table 4 suggest that the elasticity of citations with respect to a reduction in postage cost is between 0.62 and 0.73. One way to interpret this is relative to a one standard deviation reduction in the log cost, 0.36, which would imply an additional 0.22 to 0.26 citations between a dyad, relative to a sample mean of 0.18 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 1.5 log points, so lowering the postage cost for the average cost dyad to that of the lowest would lead us to expect around one additional citation over the ten-year post-reform period. Of course, these figures only apply to locations where a cited scientist was present sometime during our sample period.

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.

[35] We have fewer origin locations that destinations, so this is more stringent than clustering by destination-time. It is also more sensible, since the most likely source of correlated standard errors across citations is that they originate from the same article. We have also experimented with multidimensional clustering at the dyad, origin and destination level, finding the same levels of significance.

Table 4: Citations analysis results

DV: Number of citations between an origin-destination pair in a period						
	Robust Standard Errors			SEs clustered by origin-period		
	1	2	3	4	5	6
Log cost	-0.928***			-0.928***		
	(0.188)			(0.172)		
Log cost x post-reform	0.621**	0.718***	0.731***	0.621**	0.718***	0.731***
	(0.269)	(0.253)	(0.242)	(0.296)	(0.246)	(0.226)
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	4,524	364	220	4,524	364	220

In Columns 1-3, robust standard errors are presented in parenthesis. In Columns 4-6, standard errors are clustered by origin-period. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

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 files provide a second margin along which we can assess the impact of lower communication costs, and one that is particularly important for 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 remoteness of a location, its market access, or the presence of a railroad connect, instead of being able to include 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 letter costs, but also of distance and population. The postal reform only changes the 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

$$\Delta LMA_n = \underbrace{\ln\left(\sum_{i'\neq n} P_i^{\hat{\beta}} d_{ni'}^{-\hat{\gamma}} 1^{-\hat{\eta}}\right)}_{LMA_{n,after}} - \underbrace{\ln\left(\sum_{i'\neq n} P_i^{\hat{\beta}} d_{ni'}^{-\hat{\gamma}} c_{ni}^{-\hat{\eta}}\right)}_{LMA_{n,before}} \tag{10}$$

This value can now be computed, since P_n is the local population in n, d_{ni} is the bilateral turnpike distance and c_{ni} the postage cost between locations n and i, all observed; furthermore, the elasticities β, γ and η have been estimated in Section 4. Here we present results using our preferred estimates of the elasticities from Tables 2 and 3, which use a less parametric approach. In Appendix Tables B7 and B8, we replicate the most stringent specification for patent analysis using all twelve sets of elasticities obtained varying the estimation method, the proxy for letter access, and the number of railway stations used to compute proximity, as described in Section 4.

Equipped with a measure of the treatment size of 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 (1840-50). 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,

$$PAT_{nt} = \beta_0 POST_t + \beta_1 \left(\Delta LMA_n \times POST_t\right) + X_{nt}\Gamma + \eta_n + \epsilon_{nt} \tag{11}$$

where PAT_{nt} is the number of patents associated with post town n in period t, ΔLMA_n is defined above, $POST_t$ is an indicator for the post-reform period, X_{nt} is a set of control variables, and η_n is a set of post town fixed effects. In our preferred specification, we include controls for local population, market access, and proximity to railroad distance. Market access is defined as

$$MA_n = \left(\sum_{i'} P_i d_{ni'}^{-1}\right) \tag{12}$$

and it captures a measure of economic opportunity of location n, as proxied by its turnpike distance-weighted proximity to all population centers in Great Britain. Regression results for our patent data analysis are presented in Table 5. Column 1 looks only at the change in patents over time and shows that there is an increase in the post-reform period. In Column 2, we add in our key explanatory variable, the change in letter market access in the post period. Without including controls, we find no evidence that this is related to the number of patents produced. In Column 3, we add in a control for population. We now see evidence of a positive relationship between the change in letter market access and patenting, as well as a positive effect of population on patents. In Column 4, we add in market access. A location's market access appears to have a strong impact on patenting, and once we control for market access we see clearer evidence of a positive relationship between the change in letter market access and the number of patents produced, significant at the 95% level. In the last column, we also include a control for the average distance to the two closest railway stations. This has very little impact on the estimated effect of the change in letter market access, though it does drop the statistical significance of the estimated effect to the 90% level. Increasing the treatment from the 5^{th} to the 95^{th} percentile of exposure (about .45) increases the average number of patents by 0.26, a little more than 7% of the pre-period mean in the estimation sample. Table B7 in the Appendix reproduces the last specification of this table 5 after computing ΔLMA_n using the two-step approach for both letter access proxies and varying number of stations. Table B8 in the Appendix performs a similar robustness after computing ΔLMA_n using the joint estimation of elasticities for both letter access proxies and varying number of stations. The estimated effect of the reform is very stable across all these twelve alternative estimates. 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)). While 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

Table 5: Patent data analysis results

DV: Number of patents					
	1	2	3	4	5
Post	0.681***	0.848	-0.417	-2.352*	-2.166
	(0.039)	(0.858)	(1.153)	(1.324)	(1.351)
ΔLMA_n × Post		-0.045	0.260	0.597**	0.569*
		(0.236)	(0.297)	(0.303)	(0.295)
Ln Population			0.819	-0.603	-0.475
			(0.551)	(0.810)	(1.070)
Ln MA				6.078**	5.585
				(2.895)	(3.523)
Ln Rail Distance					0.013
					(0.047)
N	700	700	700	700	700
Post Town FE	Y	Y	Y	Y	Y
N	700	700	700	700	700

Ln Rail Distance is the log of the average distance to the two rail stations closest to the post town. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

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 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." Our results show that the "generation" of ideas is intimately linked to the "exchange" of ideas. A natural implication is that by facilitating the exchange of ideas, through proximity, cities are also accelerating the generation of new 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

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.

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.

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.

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

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

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.

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.

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.

Catalini, Christian, Fons-Rosen, Christian, & Gaule, Patrick. 2016. Did cheaper flights change the direction of science?

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

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

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

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

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.

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.

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.

General Post Office. 1859. *Fifth 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.

Gruber, Jacob W. 2006. *Owen, Sir Richard (1804-1892)*. Oxford University Press, https://doi.org/10.1093/ref:odnb/21026, 03/14/21.

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.

Juhasz, Reka, & Steinwender, Claudia. 2018 (Oct.). *Spinning the Web: Codifiability, Information Frications, and Trade*. Working Paper.

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.

Lewins, William. 1865. *Her Majesty's Mails. A History of The Post-Office, and an Industrial Account of Its Present Condition.* Second edition edn. London: Sampson Low, Son and Marston.

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

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.

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

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.

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

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

Robinson, Howard. 1948. *The British Post Office: A History.* Princeton: Princeton University Press.

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.

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

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.

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.

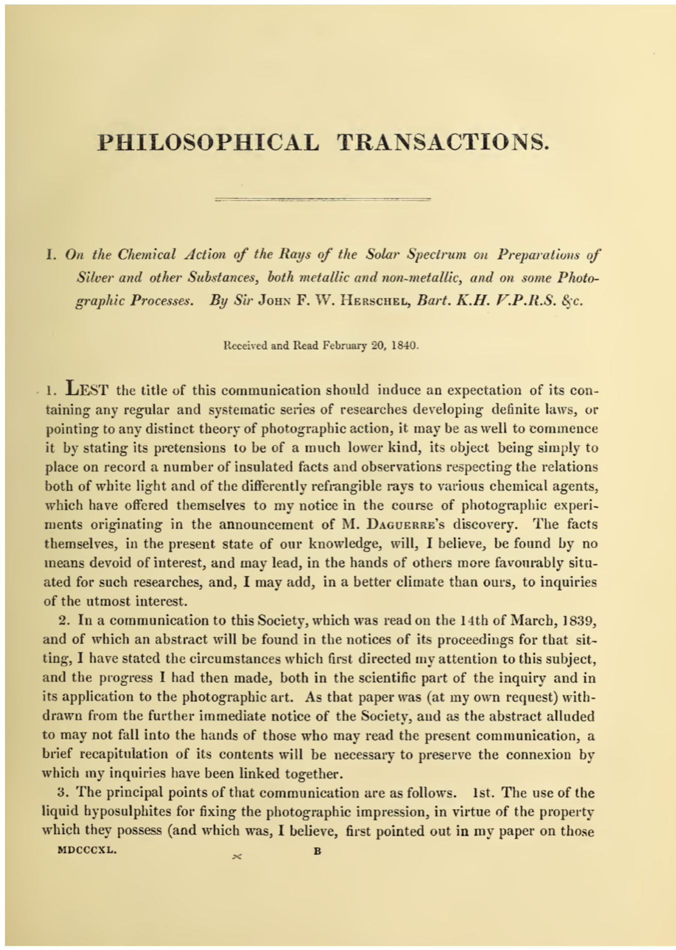
Waldinger, Fabian. 2011. Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany. *The Review of Economic Studies*, **79**(2), 838–861.

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.

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.

A Figures

Figure A1: 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.

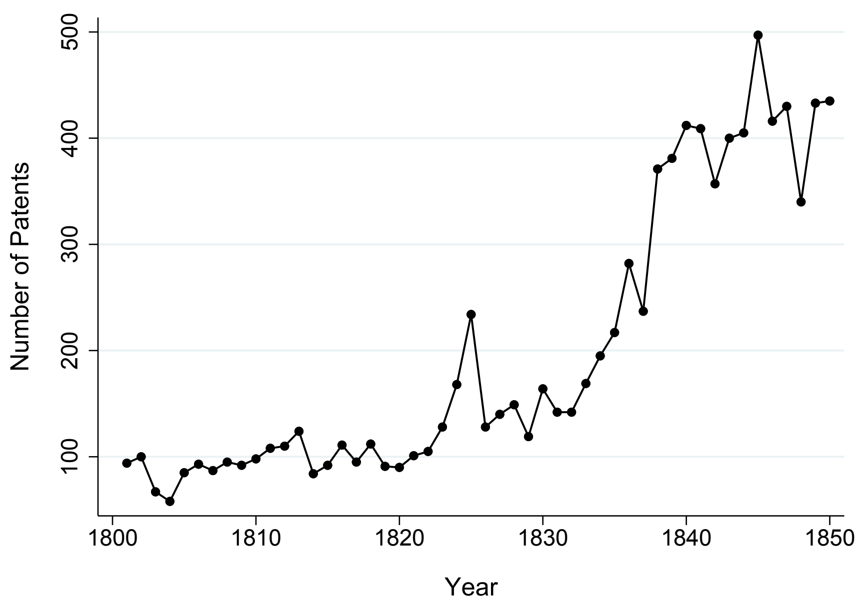
2. In a communication to this Society, which was read on the 14th of March, 1839, and of which an abstract will be found in the notices of its proceedings for that sitting, I have stated the circumstances which first directed my attention to this subject, and the progress I had then made, both in the 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 A2: Patents filed during the study period



B Tables

B.1 Summary statistics for letter flow analysis

Table B1 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. "Turnpike distance" is the distance over the turnpike network between the centroid of the Registration District of the post town and London. "Population" is the population of the Registration District where the post town belong in 1838. The variables $\widehat{LMA}_1$ and $\widehat{LMA}_1$ are the two alternative approximation to the letter market access of a post town, as described in equation 5 and 6 in the main text. To compute the "Distance to the closest station" we use the distance between post town centroids and the geolocated set of stations in 1838. Similar computations are performed for the average distance to the two closest and three closest stations.

Table B1: Summary statistics for data used in the analysis of letter flows

In Levels					
Variable	Obs	Mean	Std. Dev.	Min	Max
Letters	568	263.384	475.828	1	5660
Distance	568	124.394	72.923	12.537	327.01
Population	568	.025	.018	.002	.211
$\widehat{LMA}_1$	568	.012	.005	.005	.035
$\widehat{LMA}_2$	568	.145	.043	.064	.303
Distance to Closest Rail Station	568	51.844	41.377	.234	174.642
Average Dist. to 2 Closest Rail Stations	568	55.327	41.808	1.026	176.921
Average Dist. to 3 Closest Rail Stations	568	58.498	42.821	1.551	178.993
In Logs					
Variable	Obs	Mean	Std. Dev.	Min	Max
Ln Letters	568	4.861	1.161	0	8.641
Ln Distance	568	4.611	.707	2.529	5.79
Ln Population	568	-3.873	.546	-6.082	-1.558
Ln $\widehat{LMA}_1$	568	-4.528	.378	-5.392	-3.358
Ln $\widehat{LMA}_2$	568	-1.975	.302	-2.756	-1.196
Ln Distance to Closest Rail Station	568	3.471	1.203	-1.453	5.163
Ln Average Dist. to 2 Closest Rail Stations	568	3.636	.99	.026	5.176
Ln Average Dist. to 3 Closest Rail Stations	568	3.731	.915	.439	5.187

B.2 Robustness of letter flows regressions

Table B2 presents robustness exercises for Table 1 in the paper, where we estimated jointly the elasticity of letter flows to population, distance and postage cost. In this table, we replicate the last column of Table 1: we vary the control used for letter market access between $\widehat{LMA}_1$ in eq. 5 (Columns 1-3) and $\widehat{LMA}_2$ in eq. 6 (Columns 4-6), and the number of proximate stations between one (Columns 1 and 4), two (Columns 2 and 5) and three (Columns 3 and 6). Column 2 replicates the last column of Table 1 in the paper.

Table B2: Impact of distance on letter flows (robustness)

	DV: Log Letter Flows to London					
	1	2	3	4	5	6
Ln Postage Cost	-2.456**	-2.487**	-2.486**	-2.208**	-2.224**	-2.226**
	(1.074)	(1.077)	(1.077)	(1.077)	(1.079)	(1.078)
Ln Turnpike Distance	-0.644**	-0.631*	-0.627*	-0.612*	-0.604*	-0.603*
	(0.322)	(0.322)	(0.322)	(0.319)	(0.320)	(0.320)
Ln Population	1.187***	1.188***	1.189***	1.172***	1.173***	1.175***
	(0.097)	(0.097)	(0.097)	(0.096)	(0.096)	(0.096)
Ln $\widehat{LMA}_1$	-1.091***	-1.096***	-1.085***			
	(0.242)	(0.247)	(0.251)			
Ln $\widehat{LMA}_2$				-1.218***	-1.212***	-1.200***
				(0.282)	(0.285)	(0.288)
Ln Rail Distance (1)	-0.040			-0.024		
	(0.044)			(0.043)		
Ln Rail Distance (2)		-0.052			-0.025	
		(0.060)			(0.059)	
Ln Rail Distance (3)			-0.048			-0.018
			(0.069)			(0.067)
R^2	0.36	0.36	0.36	0.36	0.36	0.36
N	568	568	568	568	568	568
Region FE	Y	Y	Y	Y	Y	Y

Ln Rail Distance (n) is the log of the average distance to the n rail stations closest to the post town. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B3 presents robustness exercises for Table 2 in the paper, where we estimated the elasticity of letter flows to population and distance controlling for costs, in the first step of our two-step procedure. In this table, we replicate the last column of Table 2: we vary the control used for letter market access between $\widehat{LMA}_1$ in eq. 5 (Columns 1-3) and $\widehat{LMA}_2$ in eq. 6 (Columns 4-6), and the number of proximate stations between one (Columns 1 and 4), two (Columns 2 and 5) and three (Columns 3 and 6). Column 2 replicates the last column of Table 2 in the paper.

Table B3: Impact of distance and population on letter flows, controlling for cost (robustness)

	DV: Log Letter Flows to London					
	1	2	3	4	5	6
Ln Turnpike Distance	-0.866**	-0.856**	-0.845**	-0.914***	-0.899**	-0.890**
	(0.355)	(0.353)	(0.353)	(0.353)	(0.351)	(0.350)
Ln Population	1.202***	1.200***	1.203***	1.192***	1.190***	1.193***
	(0.095)	(0.094)	(0.095)	(0.093)	(0.093)	(0.093)
Ln $\widehat{LMA}_1$	-1.172***	-1.222***	-1.219***			
	(0.273)	(0.285)	(0.293)			
Ln $\widehat{LMA}_2$				-1.482***	-1.527***	-1.527***
				(0.328)	(0.339)	(0.346)
Ln Rail Distance (1)	-0.079			-0.079		
	(0.049)			(0.049)		
Ln Rail Distance (2)		-0.122*			-0.118*	
		(0.071)			(0.070)	
Ln Rail Distance (3)			-0.129			-0.126
			(0.082)			(0.081)
R^2	0.38	0.38	0.38	0.38	0.38	0.38
N	568	568	568	568	568	568
Letter cost FE	Y	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y	Y

Ln Rail Distance (n) is the log of the average distance to the n rail stations closest to the post town. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B4 presents robustness exercises for Table 3 in the paper, where we estimated the elasticity of letter flows to costs using a discontinuity design, the second step of our two-step procedure. In this table, we replicate the last column of Table 3: we vary the control used for letter market access between $\widehat{LMA}_1$ in eq. 5 (Columns 1-3) and $\widehat{LMA}_2$ in eq. 6 (Columns 4-6), and the number of proximate stations between one (Columns 1 and 4), two (Columns 2 and 5) and three (Columns 3 and 6). Column 2 replicates the last column of Table 3 in the paper.

Table B4: Impact of cost on letter flows (robustness)

	DV: Log Letter Flows to London					
	1	2	3	4	5	6
Ln Postage Cost	-1.486*	-1.572*	-1.602*	-1.442	-1.526*	-1.552*
	(0.882)	(0.872)	(0.875)	(0.877)	(0.865)	(0.865)
R^2	0.82	0.82	0.82	0.82	0.82	0.82
N	98	98	98	98	98	98
Ln Population	Y	Y	Y	Y	Y	Y
Ln Privileged Letters	Y	Y	Y	Y	Y	Y
Ln Newspaper Flows	Y	Y	Y	Y	Y	Y
Ln $\widehat{LMA}_1$	Y	Y	Y			
Ln $\widehat{LMA}_2$				Y	Y	Y
Ln Rail Distance (1)	Y			Y		
Ln Rail Distance (2)		Y			Y	
Ln Rail Distance (3)			Y			Y
Group threshold FE	Y	Y	Y	Y	Y	Y
Region FE	Y	Y	Y	Y	Y	Y

Ln Rail Distance (n) is the log of the average distance to the n rail stations closest to the post town. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

B.3 Summary statistics for the citations analysis

Table B5 presents summary statistics for the estimation sample of citations across directed post town dyads in the citation data, by period. “Pre-reform” refers to the period 1830-1839, and “Post-reform” refers to the period 140-1849. A directed dyad counts citations from location n towards scientists in location i separately from citations from location i towards scientists in location n. It also presents statistics on the log postage cost pre-reform.

Table B5: Summary statistics for the citations data

Variable	Obs	Mean	Std. Dev.	Min	Max
Citations pre-reform	2328	.18	1.212	0	30
Citations post-reform	2196	.177	1.244	0	38
Log pre-reform bilateral postage cost	2328	2.533	.36	1.099	3.114

B.4 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 140-1850. "Patents" is the count of patents filed in a location. $\Delta LMA(m, v, s)$ is the exposure to the reform calculated using either the joint estimation or the discontinuity design method, $m \in \{Joint, RD\}$ as described in Section 4; either versions one or two of our proxy for the letter market access $\widehat{LMA}_v$, $v = 1, 2$, as described in equations 5 and 6; and controlling for the average distance to one, two, or three closest rail stations, $s = 1, 2, 3$. "Population" is the population in the Registration district where the post town is located. "Market access" is the inverse of the turnpike distance-weighted 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. 12). To compute the "Distance to the closest station" we use the distance between post town centroids and the geolocated set of stations in 1838. Similar computations are performed for the average distance to the two closest and three closest stations.

Table B6: Summary statistics for the patent analysis data

Variable	Obs	Mean	Std. Dev.	Min	Max
Patents, pre-reform	350	3.649	15.565	0	194
Patents, post-reform	350	7.209	30.151	0	402
$\widehat{LMA}_1(RD,1)$	350	3.552	.13	2.972	3.848
$\widehat{LMA}_1(RD,2)$	350	3.758	.137	3.144	4.07
$\widehat{LMA}_1(RD,3)$	350	3.831	.139	3.204	4.148
$\widehat{LMA}_1(Joint,1)$	350	5.909	.207	4.959	6.379
$\widehat{LMA}_1(Joint,1)$	350	5.986	.209	5.027	6.461
$\widehat{LMA}_1(Joint,1)$	350	5.986	.209	5.027	6.46
$\widehat{LMA}_2(RD,1)$	350	3.438	.126	2.876	3.729
$\widehat{LMA}_2(RD,2)$	350	3.638	.133	3.044	3.946
$\widehat{LMA}_2(RD,3)$	350	3.702	.136	3.097	4.014
$\widehat{LMA}_2(Joint,1)$	350	5.331	.182	4.506	5.743
$\widehat{LMA}_2(Joint,2)$	350	5.372	.184	4.543	5.787
$\widehat{LMA}_2(Joint,3)$	350	5.376	.184	4.545	5.791
Population, pre-reform	350	-3.766	.54	-5.375	-1.663
Population, post-reform	350	-3.662	.576	-5.323	-1.414
Market access, pre-reform	350	-2.012	.314	-2.8	-1.243
Market access, post-reform	350	-1.875	.322	-2.684	-1.094
Ln Distance to Closest Rail Station, pre-reform	350	3.898	.949	-.651	5.199
Ln Average Distance to 2 Closest Rail Stations, pre-reform	350	4.125	.775	.946	5.254
Ln Average Distance to 3 Closest Rail Stations, pre-reform	350	4.269	.744	1.136	5.296
Ln Distance to Closest Rail Station, post-reform	350	1.913	1.274	-1.686	4.651
Ln Average Distance to 2 Closest Rail Stations, post-reform	350	2.314	.988	-1.051	4.674
Ln Average Distance to 3 Closest Rail Stations, post-reform	350	2.515	.862	-.239	4.686

B.5 Robustness of patent data analysis

Tables B7 and B8 present robust for the last column of the patent analysis in Table 5 in the paper, where we vary the way we compute the treatment size across our twelve alternative approaches. In particular, Table B7 uses treatment estimates computed using the two-step, discontinuity design approach, while Table B8 uses treatment estimates from the joint elasticity estimation approach, as described in Section 4. Within each Table we vary the control used for letter market access between $\widehat{LMA_1}$ in eq. 5 (Columns 1-3) and $\widehat{LMA_2}$ in eq. 6 (Columns 4-6), and the number of proximate stations between one (Columns 1 and 4), two (Columns 2 and 5) and three (Columns 3 and 6). Column 2 replicates the last column of Table B7 in the paper.

Table B7: Patent data analysis results (Robustness for two-step estimation of elasticities)

DV: Number of patents						
Proxy for LMA in regression letters	$\widehat{LMA_1}$	$\widehat{LMA_1}$	$\widehat{LMA_1}$	$\widehat{LMA_2}$	$\widehat{LMA_2}$	$\widehat{LMA_2}$
Number of proximate stations	1	2	3	1	2	3
Post	-2.295*	-2.166	-2.076	-2.389*	-2.256*	-2.174
	(1.365)	(1.351)	(1.388)	(1.347)	(1.333)	(1.368)
$\Delta LMA_n \times Post$	0.622*	0.569*	0.544*	0.663**	0.606**	0.582*
	(0.318)	(0.295)	(0.293)	(0.323)	(0.299)	(0.298)
Ln Population	-0.573	-0.475	-0.422	-0.585	-0.488	-0.439
	(1.005)	(1.070)	(1.122)	(1.003)	(1.068)	(1.120)
Ln MA	5.948*	5.585	5.380	6.128*	5.752	5.559
	(3.430)	(3.523)	(3.587)	(3.426)	(3.515)	(3.578)
Ln Rail Distance	0.003	0.013	0.021	0.002	0.012	0.019
	(0.034)	(0.047)	(0.065)	(0.034)	(0.047)	(0.065)
N	700	700	700	700	700	700
Post Town FE	Y	Y	Y	Y	Y	Y

Ln Rail Distance is the log of the average distance to the two rail stations closest to the post town. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B8: Patent data analysis results (Robustness for joint estimation of elasticities)

DV: Number of patents						
Proxy for LMA in regression letters	$\widehat{LMA_1}$	$\widehat{LMA_1}$	$\widehat{LMA_1}$	$\widehat{LMA_2}$	$\widehat{LMA_2}$	$\widehat{LMA_2}$
Number of proximate stations	1	2	3	1	2	3
Post	-2.185	-2.031	-1.944	-2.157	-2.002	-1.917
	(1.416)	(1.412)	(1.453)	(1.444)	(1.441)	(1.479)
$\Delta LMA_n \times Post$	0.364*	0.344*	0.335*	0.402*	0.382*	0.372
	(0.201)	(0.196)	(0.199)	(0.229)	(0.225)	(0.227)
Ln Population	-0.520	-0.421	-0.369	-0.511	-0.409	-0.355
	(1.010)	(1.075)	(1.127)	(1.011)	(1.077)	(1.128)
$Ln\ MA$	5.523	5.139	4.951	5.380	4.999	4.817
	(3.435)	(3.532)	(3.598)	(3.428)	(3.529)	(3.592)
Ln Rail Distance	0.005	0.016	0.024	0.006	0.017	0.025
	(0.034)	(0.047)	(0.065)	(0.034)	(0.048)	(0.065)
N	700	700	700	700	700	700
Post Town FE	Y	Y	Y	Y	Y	Y

Ln Rail Distance is the log of the average distance to the two rail stations closest to the post town. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

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.

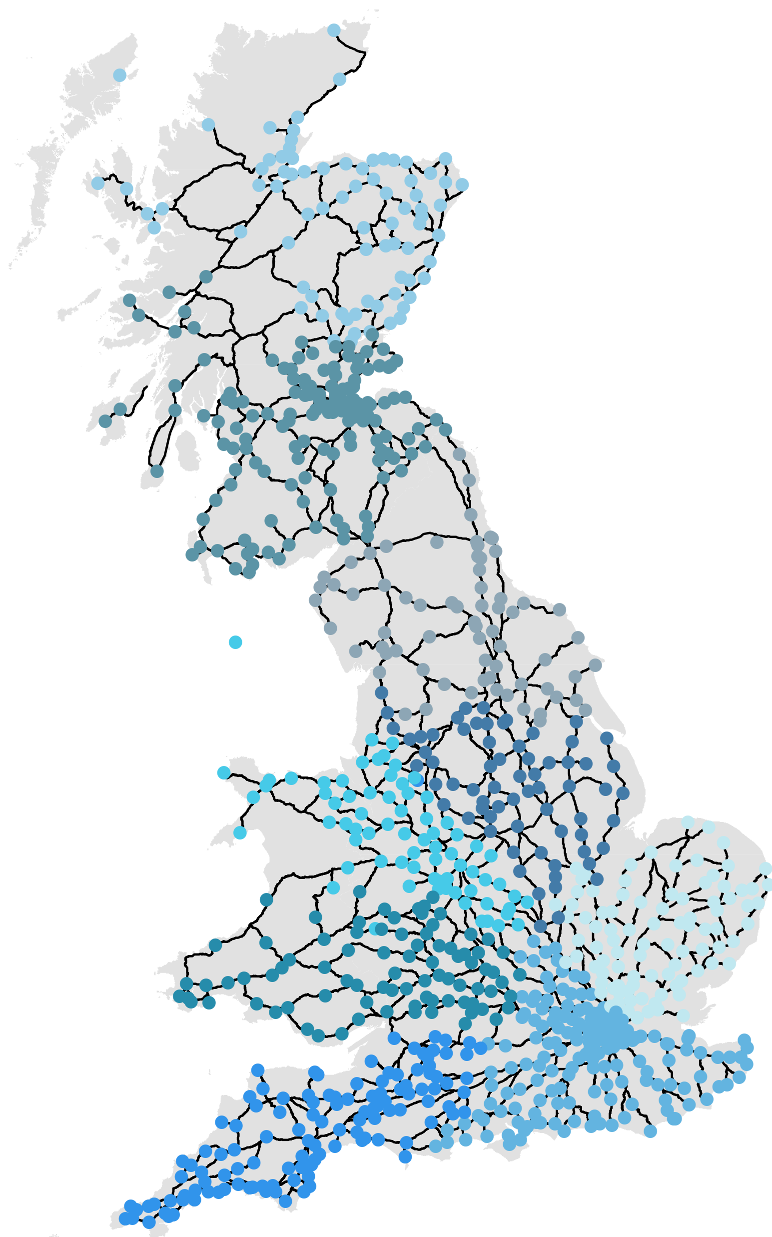
Post Towns Our list of post towns in 1838 is taken from a publication by the *Select Committee on Postage*. It lists 862 post towns and over 1,600 subpost towns in England, Scotland or Wales that are nested in the following eight post districts:

1. Home Districts
2. Eastern Districts
3. Western Districts
4. Midlands District
5. Northern District
6. North Wales district
7. South Wales District
8. Northern Scotland District
9. Southern Scotland District

To geolocate the post and subpost towns, we first make use 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 subpost towns. Figure A3 shows the post town locations with different colors indicating the eight post districts listed above. 3 shows a map of the subpost town locations.

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 A4 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

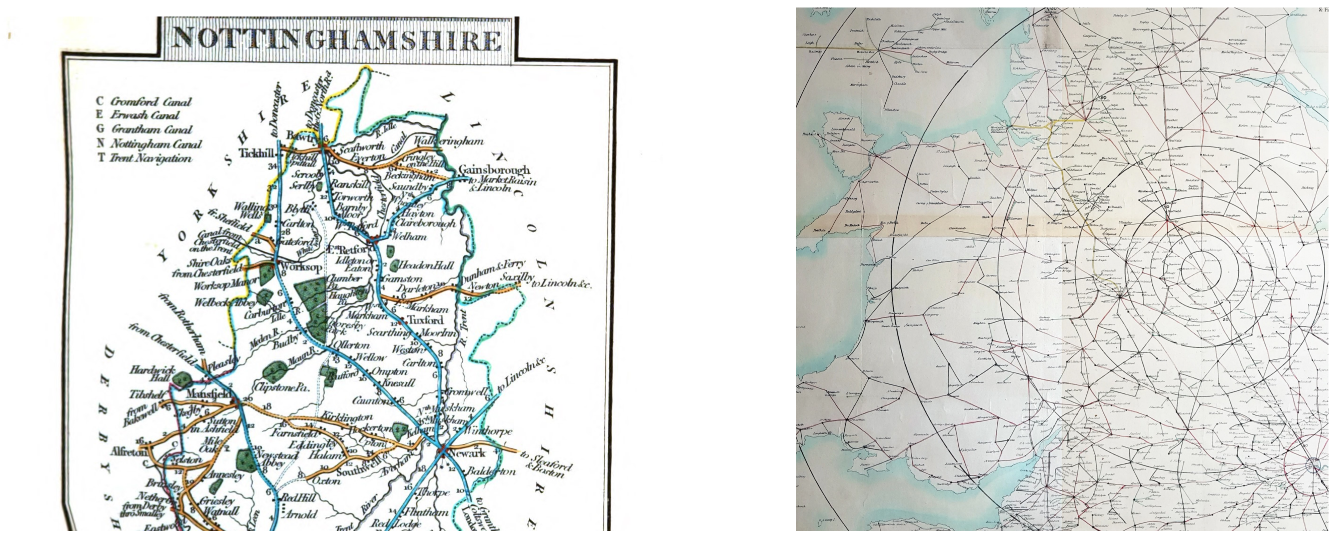
Figure A3: Post towns and post roads



This figure shows the network of post roads connecting post towns in 1838. Different colors of post towns refer to the eight postal districts in England, Scotland and Wales.

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 an assign zero-distance to these connector bits. The rationale is that we do not consider costs of sending letters within post towns and therefore assume that every post town is located along the post road.

Figure A4: Historical Maps



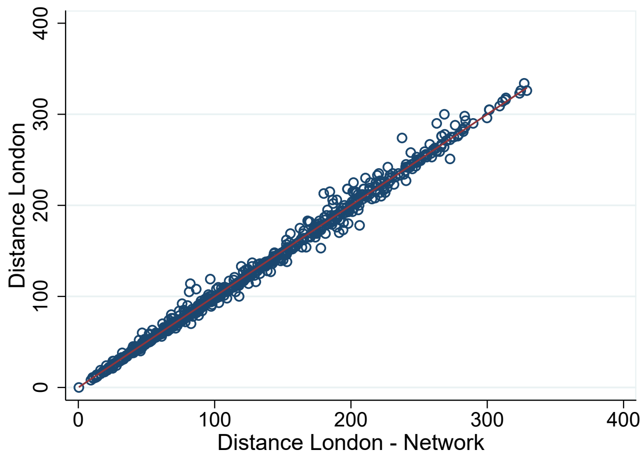
(a) Cary 1828

(b) Basire 1838

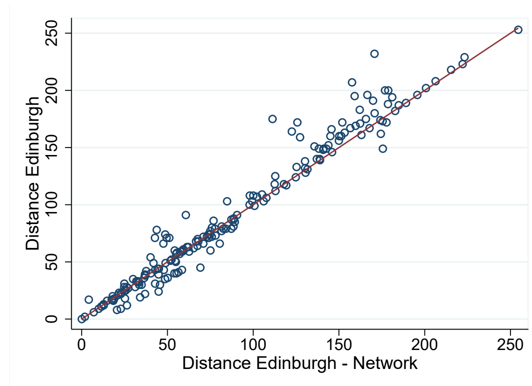
This figure shows excerpts from Cary (1828) and Basire (1838) maps which were digitized to compute postal network routes.

To test how well our postal network approximates historical travel distances, we make us 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 A5 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. These differences are of little concern to our analysis since we do not use Scotland in the analysis of patent data and we only observe very few authors located in the Highlands.

Figure A5: Comparison



(a) Distance to London



(b) Distance to Edinburgh

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

Postage To derive the costs of sending a letter before the introduction of the flat postage, we multiply the distance between all post-towns pairs with the distance-dependent postage for a single sheet letter displayed in the first column of Table B9.

Connections We are ultimately interested in the costs of sending a letter between locations of scientists or inventors. To link scientists' locations to the postal network, we assign each author to the nearest subpost town or posttown and use a post office publication that links each subpost towns hierarchically to a post town to calculate the costs of sending a letter between post offices. We neglect the costs of sending a letter between subpost towns and post towns and post offices since they were typically a flat fee of a penny that would likely apply to all flows. In the worst case, this would imply that we are understating the change in communication costs. Since we do not have address-specific patent data but only observe the applicants home town, we geolocate all home town locations and then link each patentee location to the nearest subpost town or post town.

Table B9: Costs of sending a letter in 1838

Mileage	Single	Double	Treble	Ounce	+ea. 1/4oz
Up to 8 miles	2d	4d	6d	8d	+ 2d
8-15 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 figure shows how postage rates increase with distance. Single, double, treble refers to the number of sheets. d: penny; s: schilling; $12d = 1s$.

D Construction of the Citation Dataset

D.1 Summary Statistics

443 articles appeared in the *Philosophical Transactions of the Royal Society of London* between 1830 and 1849. Table B10 displays basic descriptive statistics for this population at the year, period, and aggregate level.

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.

Contrasting with today's practices, it was not uncommon for an individual author to publish multiple articles in the same issue, which explains why the number of articles exceeds the number of authors in most years. Co-authorship was rare. Solely the first article in the 1844 issue lists two authors. However, more articles represent joint work. These articles typically state a single author but include notes by other scientists that often extend to several pages. Subsection D.2 explains the handling of the different forms of co-authorship. Table B10 only takes official co-authorship into account, meaning that only the first article in 1844 is considered as coauthored.

Three articles in the sample contain sizable parts that were written by scientists other than the stated authors. These three articles and the mentioned 1844 article were split up to attribute each forward citation to its origin. One of the three articles was published in the pre-reform period, the other two

Table B10:
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

appeared in the post-reform period, leading to article counts of 252 in the pre- and 195 in the post-reform period and a total count of 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 B11 zooms in on these articles. The coauthored articles are already split up in table B11. Table B12 displays summary statistics for all citations in these articles.

The citation counts in table B12 are restricted in the following way: total citations: all citations we have entered. Regular citations: citations excluding references to military and other personnel who conducted measurements and are only listed in the tables that display these measurements. Citations to living scientists: citations to scientists who were alive in the publication year. Citations to geolocated scientists: citations for which we were able to find the locations of the author and the cited scientist. Citations to scientists in Great Britain: citations that connect two scientists who resided in England, Wales, or Scotland in the publication year of the article which contains the citation. The bases of the percent shares are the following: regular citations: total citations, citations to living scientists: regular citations, citations to geolocated scientists: citations to living scientists, citations to scientists in Great Britain: citations to geolocated scientists.

D.2 Correspondence and Co-authorship

D.2.1 Correspondence

Authors who were not affiliated with the *Royal Society of London* typically did not directly communicate their findings to the editors of the *Philosophical Transactions of the Royal Society of London*. In these cases, the fellows of the *Royal Society of London* intermediated between the unaffiliated authors and the editors. This intermediation is defined as correspondence. 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. Most correspondents are also regularly cited in these articles. We recorded whenever a correspondent was not also regularly cited in an article to be able to exclude correspondence links in the robustness checks of our analysis.

Table B11:
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

Table B12:
Citations by Authors who lived in Great Britain in the Publication Year

Period	**total citations**	**regular citations**		**to living scientists**		**to geolocated scientists**		**to scientists in Great Britain**	
	count	count	share	count	share	count	share	count	share
1830	147	146	99.3	101	69.2	100	99.0	60	60.0
1831	170	166	97.6	91	54.8	88	96.7	44	50.0
1832	268	253	94.4	163	64.4	154	94.5	80	51.9
1833	405	404	99.8	245	60.6	216	88.2	98	45.4
1834	209	209	100.0	118	56.5	115	97.5	56	48.7
1835	201	201	100.0	135	67.2	132	97.8	52	39.4
1836	291	218	74.9	144	66.1	143	99.3	75	52.4
1837	202	200	99.0	120	60.0	120	100.0	72	60.0
1838	175	175	100.0	90	51.4	89	98.9	35	39.3
1839	209	209	100.0	152	72.7	151	99.3	73	48.3
Pre period	2,277	2,181	95.8	1,359	62.3	1,308	96.2	645	49.3
1840	255	255	100.0	192	75.3	190	99.0	67	35.3
1841	201	199	99.0	142	71.4	140	98.6	64	45.7
1842	147	146	99.3	108	74.0	108	100.0	57	52.8
1843	146	145	99.3	114	78.6	114	100.0	50	43.9
1844	144	144	100.0	95	66.0	95	100.0	51	53.7
1845	78	77	98.7	61	79.2	60	98.4	28	46.7
1846	310	309	99.7	212	68.6	209	98.6	91	43.5
1847	102	101	99.0	53	52.5	53	100.0	25	47.2
1848	128	128	100.0	97	75.8	94	96.9	53	56.4
1849	219	219	100.0	178	81.3	169	94.9	88	52.1
Post period	1,730	1,723	99.6	1,252	72.7	1,232	98.4	574	46.6
Total	4,007	3,904	97.4	2,611	66.9	2,540	97.3	1,219	48.0

D.2.2 Co-authorship

Two types of co-authorship occur. One article lists two authors. Three other articles include parts that were written by scientists other than the stated authors. The officially co-authored article was treated as if each coauthor had individually written the article with the exception that a non-standard citation to the coauthor was added. One of the authors is also cited in the article, so that the value was changed back to a regular citation in one of the two separated articles. Accounting for co-authorship in this way increased the article count by one, the page count by nineteen, and resulted in the double-counting of eight citations.

The three articles which contain contributions by scientists other than the stated authors were split up. The parts that were not written by the stated authors were treated as separate articles. The pages and citations were allocated accordingly. This increased the article count by three, did not enlarge the page count, and resulted in a modest increase in citations in cases where the same scientists were cited in both separated parts of the initial articles. Co-authorship occurs once in the pre period and three times in the post period. Analogously to correspondence, if a co-author is also regularly cited in a split-up coauthored article, we count this as a regular citation but keep track of this to exclude such citations in the robustness checks.

D.2.3 Other Non-standard References

15 articles do not contain any citations. One article header states the name of the author's supervisor. This reference was recorded and excluded in the analysis.

D.3 An Algorithm for Identifying the Cited Scientists

1. Search for the cited scientist's last name in the spreadsheet `scientists`

   1.1. If you don't find a scientist with the same last name, go to 2.

   1.2. If you find a scientist with the same last name, conduct a quick check whether the fields provided in f1 to f4 or the perma.cc link line up with the content around the citation. For example, is the author writing about the cited scientist's contribution to photography in the article, but the scientist with the same last name in the spreadsheet is supposed to be a medical doctor? Repeat this step if there are multiple scientists with the same last name.

      1.2.1. If there is only one scientist listed in the spreadsheet `scientists` whose fields do not substantially differ from the content around the citation, take note of the scientist's first name or initials and go to 3.

1.2.2. If there are multiple scientists in the spreadsheet `scientists` whose fields do not substantially differ from the content around the citation, check whether any of the scientists listed in the spreadsheet `scientists` were fellows of the *Royal Society* by searching for the scientist's last name in the fellow directory: https://royalsociety.org/fellows/fellows-directory/.

1.2.2.1. If one or more of the scientists in the spreadsheet were fellows, take note of the first name(s) of these scientists and go to 3.

1.2.2.2. If none of the scientists was a fellow, go to 2.

1.2.3. If the content around the citation and the entries in f1 to f4 or the perma.cc link differ substantially, go to 2.

2. Search for the scientist's last name in the fellow directory: https://royalsociety.org/fellows/fellows-directory/.

2.1. If you don't find a fellow with the same last name, go to 4.

2.2. If you find a fellow with the same last name, check whether the fields indicated in the entry of the fellow directory line up with the content around the citation in the article. Repeat this step if there are multiple scientists with the same last name.

2.2.1. If there isn't a clear indication that the fields and content around the citation differ substantially, take note of the scientist's first name or initials and go to 3.

2.2.2. If the content around the citation and the stated fields differ substantially, go to 4.

3. Search for the scientist's last name in the *Catalogue of Scientific Papers*. Go to the scientist who has the same first name (initials) and search his/her publications. Are his/her publications in the same research area as the content around the citation? Repeat this step if there are multiple scientists with the same first name (initial) and last name.

3.1. If the research areas add up, go to 6.

3.2. If the research areas don't add up, go to 4.

4. Look up the scientist's last name in ODNB: http://researchguides.library.vanderbilt.edu/c.php?g=69225&p=445618. Check whether there are entries in ODNB that match the scientist's name, field of research, and time he/she was alive. Take note of potential matches. Search for the scientist's last name in the *Catalogue of Scientific Papers*. If 3 got you here, this means extending the search to scientists with different first names or initials.

4.1. If there are some potential matches from the search in ODNB, check the first names or initials of the potential matches first. Are any of the scientist's publications in the same research area as the content around the citation?

4.1.1. If a single or multiple scientists with the same last name have matching research fields, go to 11.

4.1.2. If there aren't any scientists with matching research fields go to 5.

4.2. If there are no potential matches from the search in ODNB, go through all scientists who have the same last name and search their publications. Are any of the scientists' publications in the same research area as the content around the citation?

4.2.1. If a single or multiple scientists with the same last name have matching research fields, go to 11.

4.2.2. If there aren't any scientists with matching research fields go to 5.

5. Conduct some quick Google searches. Search for the name of the author of the article and the name of the cited scientist e.g. "Faraday Jenkin". Search for the cited scientist and any words around the citation that indicate a specific research field, e.g. "Jenkin photography". Search for the last name and "scientist".

5.1. If this provides you with convincing information and credible sources (Wikipedia is fine) that enable you to identify the cited scientist, go to 10.

5.2. If this does not provide you with convincing information on the identity of the cited scientist, go to 12.

6. Conduct a quick check to verify whether the scientist cited in the article and the scientist with the same last and first name in the `scientists` spreadsheet are really the same person. Check the perma.cc links for this person, check Wikipedia, most importantly check ODNB.

6.1. If there is convincing evidence that you have identified the scientist, go to 8.

6.2. If you are not convinced that you have identified the scientist, go to 7.

7. Look up the scientist's name in ODNB. Check whether there is an(other) entry in ODNB that matches the scientist's name, field of research, and time he/she was alive.

7.1. If you are convinced that the scientist already listed in the spreadsheet `scientists` is the one cited in the article, go to 8.

7.2. If you are convinced that another scientist with an entry in ODNB is the one cited in the article, go to 9

7.3. If there is no such entry, or you are not convinced either way, go to 4.

8. Go to the spreadsheet `scientists`, hold down command if using Mac OS, crtl if Windows, and select the cell with the `sc_id` of the identified scientist, then click on the cell that contains the scientist's first and last name in the column `fn_ln`. Press C to copy both the `sc_id` and the `fn_ln`. Go to the spreadsheet `citations`, right-click on the next free cell in `cited_id` and select "paste special" -¿ "values". This copies the scientist's id and first and last name. It is important to past values. Simply pasting what you have copied will copy the formula for `sc_id`, which will result in a wrong `sc_id` in the spreadsheet `citations`.

9. Quickly check whether the scientist you have identified could possibly have published the articles listed in the *Catalogue of Scientific Papers* that you just looked at earlier.

   9.1. If the scientist passes this test, go to 10.

   9.2. If the scientist does not pass this test, go to 12.

10. Add the first and last name of the scientist you just identified to the columns `fns` and `sn` of the spreadsheet `scientists`. `sc_id`, `alpha_id`, and `fn_ln` will get filled automatically. Add as much information as you have found to the columns, but don't invest a ton of time for minor things. Once you have entered the year of birth and year of death, `sc_id` will be determined. Go to 8.

11. Conduct some Google searches. Search for the scientist's last name and "scientist". Search the first and last name of the different matches. Search for the name of the author of the article and the name of the cited scientist e.g. "Faraday Jenkin". Search for the cited scientist and any words around the citation that indicate a specific research field, e.g. "Jenkin photography".

   11.1. If there are credible sources (Wikipedia is fine) that enable you to identify the scientist based on the criteria tested above, go to 10.

   11.2. If the Google search does not provide sufficient insights on the identity of the scientist, go to 12.

12. Look up the scientist's name in Clifton (1995). Check whether there is an entry that matches the scientist's (or instrument maker's) name, field of research, and time he/she was alive.

   12.1. If there is such an entry, go to 10.

12.2. If there is no such entry, go to 13.

13. Go to the spreadsheet `scientists`. Enter as much of the author's name as you have. Start the entry in `fns` with "NOT FOUND" followed by any information on the forenames. Go to the column `notes` and provide any information you have found about the scientist. Indicate how credible you think this information is and provide a link to the sources. I have access to basically all existing census records, registers, etc. and this information can help me to eventually identify the scientist.

D.4 Location Data Entry Instructions

This project involves entering location data for 19th-century scientists from pdf files of short biographic articles to an Excel file. Each pdf contains biographic information on one particular scientist. We are interested in each scientist's modal location in the years 1830 to 1849. Of particular interest are the modal locations in the years a scientist was cited.

The provided Excel file identifies the scientist by year combinations for which location data should be entered. The required location information is extracted from the biographic articles, other sources I make available, and additional sources you may find. For each scientist by year combination of interest, three pieces of information should be added to the Excel sheet: The scientist's institution, address, and geographic coordinates. The example in section D.4.1 elucidates the entry of these items.

The remainder of these guidelines is organized as follows. Section D.4.2, D.4.3, and D.4.4 explain the entry of institution, address, and coordinate data, respectively. Section D.4.5 provides a recipe for navigating the data sources in our database and shows how additional sources can be found, used, and documented. Section D.4.6 describes the operation of the data entry in Excel.

D.4.1 Example

The highlighted location information in the following three text snippets from a biographic article on Richard Owen (1804-1892)[36] can be used to enter location data for the years 1828-1832.

> 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

[36] see Gruber (2006).

The following table depicts five rows in the locations tab of the Excel file prior to the data entry.
BEFORE:

sc_id	alpha_id	fn_ln	year	location		coor
5528178	OwenR	Richard Owen	1828			
5528178	OwenR	Richard Owen	1829			
5528178	OwenR	Richard Owen	1830			
5528178	OwenR	Richard Owen	1831			
5528178	OwenR	Richard Owen	1832			

The table below shows the same rows after the data entry.
AFTER:

sc_id	alpha_id	fn_ln	year	location		coor
5528178	OwenR	Richard Owen	1828	Hunterian Museum, Royal College of Surgeons, London, England		51.515441,-0.116844
5528178	OwenR	Richard Owen	1829	Hunterian Museum, Royal College of Surgeons, London, England		51.515441,-0.116844
5528178	OwenR	Richard Owen	1830	Hunterian Museum, Royal College of Surgeons, London, England		51.515441,-0.116844
5528178	OwenR	Richard Owen	1831	Hunterian Museum, Royal College of Surgeons, London, England		51.515441,-0.116844
5528178	OwenR	Richard Owen	1832	Hunterian Museum, Royal College of Surgeons, London, England		51.515441,-0.116844

In this example the institution is *Hunterian Museum, Royal College of Surgeons* and the address is *London, England*.

D.4.2 Entering Institution Data

Many scientists were affiliated with multiple institutions. Data should only be entered for the main institution; this is the organization in which the scientist spent most of his/her research time in a given year. Typically, the main institution will be the scientist's employer or place of work. The main institution of a research-active business owner will be her store and the main institution of a wealthy lord will be his residence.

Scientists were often members or fellows of a variety of societies and associations. These organizations worked in a similar way as book or sports clubs work today. If a society or association had a physical location, most of its members would not have been present at this location for most of the year. This means the primary institution will typically not be a society or association. Exceptions are full-time positions to run a society or association. Googling the name of an institution helps to figure out the institution type and address.

D.4.3 Entering Address Data

Level of Detail

- In general, if an address is in the UK, always state the specific country, e. g. *England*, *Wales*, or *Scotland*, not just *United Kingdom.*

- If the institution is still in existence today, like the *Hunterian Museum* in the previous example, and its location constitutes a landmark such as the *Royal College of Surgeons*, it will suffice to enter the name of the institution, the city and the country.

- If the institution no longer exists or its location is not commonly known, any provided information about the address should be entered, e. g. *Bristol Institution, Queens Road, Bristol, England.*

- If the source provides a commercial or residential street address in a comparably large city and the location is neither determined by the institution nor generally well-known, the street name and number should be entered, e. g. *13 Hillsbridge Place, now Clarence Road, Bristol, England.*

- In the case of a small village, town, or part of town, it won't be necessary to enter the street address. However, to distinguish between towns with similar names, the name of the district or county should be entered if provided, e. g. *Great Malvern, Worcestershire, England.*

Multiple Addresses

- If a biographic article lists multiple addresses for a given year, the address at which the scientist spent the majority of his/her (research) time in that year should be entered.

- If the article states both a business and a residential address for a given year, you should enter the business address if the business is related to the scientist's field and the residential address otherwise.

- Suppose you wanted to send the scientist a letter. What address would you use to make sure that the scientist will receive and read your letter as soon as possible? The answer to this question is the address you should enter.

No Address

- The biographic articles will not provide location information for each year and scientist combination.

- If a biographic article is comparably extensive and detailed, we will assume that the scientist stayed at the previously stated location for up to three additional years or until the article mentions a new location.

- However, if there is any indication that the scientist might not have continued to stay at the previously stated location, the cells that correspond to the years without location information should be left blank.

- If a biographic article is comparably short or lacks detail, the cells that correspond to the years without location information should be left blank.

Travel

- Travel only matters if a journey or the stay at a particular travel destination represents a scientist's modal location for a given year; that is the location at which the scientist spent the most days in that year. Here is an example: *"In 1839, she spent three months in Paris, four in Rome, and two in Milan."* The modal location in this case is *Rome, Italy* since $\frac{4}{12}$ in Rome ¿ $\frac{3}{12}$ in Paris $\geq \frac{3}{12}$ unknown ¿ $\frac{2}{12}$ in Milan.

- Entering address data for domestic or foreign travel destinations won't be necessary if the scientist's modal location for a given year is his/her regular address in Britain. The regular address should be entered in this case.

- If a scientist's modal location in a given year is a travel destination, the location information should in general be entered like a regular address as explained in the previous subsections.

- If a scientist's modal location in a given year is a city outside of Britain and no street name is provided, enter the name of the city and the country. If the city population at the time exceeded 500,000, it's advantageous to also include the district. If the city's name or country changed, state the location like this: *Pera, Constantinople, Ottoman Empire, now Beyoglu, Istanbul, Turkey.*

- If a scientist's modal location in a given year is a region with a population center outside of Britain, enter the name of the region, the country, and the population center like this: *Travel to Lombardy, coordinates for Milan, Italy.*

- If a scientist's modal location in a given year is a region without a clear population center or with multiple population centers, enter the name of the population center closest to the geographical center and add the region and country in the same way as in the example above. If the region is a remote rural area, enter the name of the region and the country and add coordinates for the geographical center of the region.

- If a scientist spent the majority of a given year traveling and the information about the travel destinations is imprecise, enter the location data of the population center closest to the geographical center and the travel information. E. g. if a scientist travelled to Germany, Austria, and Hungary, enter *Travel to Germany, Austria, and Hungary, coordinates for Vienna, Austria.*

- If a scientist spent the majority of a given year traveling and visited numerous destinations, it can be helpful to group several small, close-by geographic areas into regions to determine the modal location. The following constructed example of a French geologist who explored glaciers in the Alps illustrates how these rare, but challenging cases can be addressed.

  EXAMPLE:
  He left Paris in late March and briefly explored the Ötztal glaciers in the Austrian Alps before moving on to study the Vorab, Grindelwald, and Rhone glaciers in the Swiss Alps. After completing his research there, he checked in at Wengen and made several trips to the Aletsch and Fiescher glaciers in the summer months. He completed his book on Alpine glaciers in early September but decided to extend his stay in the Bernese Oberland to enjoy some leisure hiking. He left the region and met with two old friends in Geneva on September 29. The group decided to deviate from the initial plan of returning directly to Paris and instead went on a trip to Montpellier. They arrived in Paris on November 15.

  In this case the modal location is *Swiss Alps* and a possible location entry is *Swiss Alps, coordinates for Wengen, Lauterbrunnen, Switzerland.*

  Notice that Wengen itself would not be the modal location and that a case can be made to use Paris as the modal location for this year. Looking up the glaciers and locations on Wikipedia and Google Maps and keeping track of the lengths of the scientist's stays in each region tells us the following: The Aletsch and Fiescher glaciers are located in the Bernese Alps. The Bernese Alps are part of the Swiss Alps. Wengen is located in the Bernese Oberland, which in the broader sense is also part of the Bernese Alps. Assuming that the scientist didn't spent more than a month in the Austrian Alps and taking "summer months" literally as June, July, and August, the aggregated region *Swiss Alps* includes:

  - Vorab, Grindelwald, and Rhone glaciers in May
  - Wengen, Aletsch and Fiescher glaciers in June, July, and August
  - Bernese Oberland in September

  Swiss Alps is the modal region in the given year because 5 months in the Swiss Alps ¿ 4.5 months in Paris, France (Jan-Mar, Nov 15-Dec) ¿ 1.5 months in Montpellier, France ¿ 1 month in Öztal, Austria. Coordinates should be entered for the modal location within the modal region. In this case that location is *Wengen, Lauterbrunnen, Switzerland*.

D.4.4 Entering Coordinates

The entry of geographic coordinates is explained here: `https://perma.cc/FT8J-4VRY`. Please stick to the format that is used in the column `coor` in the example in section D.4.1 and zoom in as much as possible.

D.4.5 Data Sources

If a scientist's biographic article doesn't contain all the required location information or no biographic article is available for the scientist, we can check the other sources in our database, search for additional sources, or use historical dictionaries. The following subsections and individual paragraphs describe different search procedures and are ordered from lowest to highest involved search effort.

Navigating the Data Sources in the Database

1. Look up the scientist in our database by searching for his/her id in the scientists tab. The columns K to N in the scientists tab provide perma.cc links for most scientists. These links lead to additional sources that often provide valuable location information.

2. If the columns K to N in the scientists tab don't contain any perma.cc links, logging in to perma.cc and searching for the scientist id is worth a try since some of the links weren't copied over to the Excel file. Please copy over the perma.cc links to the Excel file whenever you come across such a case.

3. If there also aren't any links on perma.cc, try to find the other sources that are provided in the columns K to N. If you cannot find these sources or the columns K to N don't provide any helpful and accessible information, you should check whether the identity of the respective scientist can be pinned down by his/her full name, date of birth, date of death, and/or the notes in column J.

4. If this information allows you to pin down the scientist's identity, you should follow the instructions in subsections D.4.5, D.4.5, and D.4.5 for finding, using, and documenting additional sources, respectively.

5. If the information in the scientists tab doesn't allow you to pin down the identify of the scientist, you should send me an email with the scientist's id so that I can look into this.

Finding Additional Sources

- Searching for the scientist's name and the year for which we need location data on Google Books often leads to sources that contain the required location information.
- Performing the following search routine on Google and Google Books is generally quite helpful.
  - Google the scientist's name.
  - Google the scientist's name and the year for which we need location data.
  - Extend your Google searches to adjacent years.
  - Google combinations of the scientist's name and presumed institutions or locations.
- The references in the provided biographic articles typically include useful sources. Some of the referenced sources are full biographies on the respective scientist. Many of these biographies date back far enough to no longer be protected by copyright so that you can find a pdf of the biography by simply googling the author and the title of the biography.
- Sometimes scientists spent longer periods of time with other prominent people. Biographic information on these individuals then can be used to find the location of the scientists.

Using Historical Dictionaries

All sources listed in this subsection are available as pdfs. Most of these pdfs are searchable but in contrast to digitally created original content, the text that is being searched differs from the text you see. The pdf files combine scans of the historical sources with text that was extracted from these scans via optical character recognition (OCR). The quality of the extracted text varies but is often limited. You should try to search for the scientist's name in the dictionaries. If you don't find the scientist, you have to resort to manual alphabetical search. Any source mentioned in this subsection can be directly accessed by clicking on the name of the source.

- Poggendorff's (1863) dictionary of scientists concisely states the vita of each included scientist. Location information can be extracted directly from the institutions that are listed in a scientist's vita. The dictionary is comprehensive. The coverage is particularly good for scientists from Continental Europe. The dictionary's disadvantage is that it is written in German, which makes using this source more time-consuming for you. Since the text was OCR-ed, the usefulness of Google Translate or other translation tools is limited. Copying a scientist's vita over to Google Translate and using this German-English dictionary should allow you to take advantage of this source without wasting time.

- Several directories and lists cover specific professions. You have access to 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. Directories like these exist for most professions. Please reach out if you cannot find the location of an astronomer, engineer, physician, or member of a different profession or learned society, so that I can provide you with the respective directory or list.

- The *Catalogue of Scientific Papers* is one of the two last resorts for finding location data. The catalogue lists virtually all scientific articles that were published between 1800 and 1860. The process below shows how you should use the catalogue. Stop this process whenever you have obtained the required location information.

  1. Check whether the scientist published an article in or around the year of interest.
  2. Write down the article title, year, journal abbreviation, volume, issue, and page numbers.
  3. Perform Google and Google Book searches with the scientists' name and the article title.
  4. Open the abbreviation index and use the journal abbreviation to obtain the journal name from the index.
  5. Do a Google search with the journal name and volume to get to a website such as biodiversitylibrary.org, archive.org, or hathitrust.org that provides access to the journal.
  6. Use the issue and page number to find the article. Check whether the institution or location of the scientist is stated on the first or last page of the article.

- If the combined information you obtained from conducting the searches described in this section doesn't suffice to enter the required locations, you should send me an email with the scientist's id. I will then perform a final search in some other records such as the British censuses.

Documenting Additional Sources

- Use perma.cc to document location information that you obtained from webpages. Start the title of each new permanent link with the correct scientist id. The scientist id is the number that is shown at the beginning of the file name of each biographic article. After entering the scientist id, you should add the scientist's name to the link title. Add all links to the folder `Penny Post`.

- Use the folder `documenting_other_sources` to save additional pdfs, ebooks, or screenshots. Give these files the same name as the initial biographic article but add `_01` for the first additional source, `_02` for the second, and so on to the file name.

D.4.6 Data Entry

The assignment tab in your version of the Excel file contains a list of scientists. Click on the blue boxes in column F to go directly to scientist-year combinations for which you should enter data. After entering location data for a combination, record the scientist's annual modal locations in other years if that does not require much additional time.

When you have completed the location data entry for a scientist, click on the assignment tab to go back to your list. Once you have entered the required location data for a scientist in your list, the blue box in that row will disappear. Clicking on the next blue box in the list will take you to the next scientist by year combination for which you should enter location data.

We extract most of the required location information from short biographic articles. The pdf files of these biographic articles are stored in the regularly updated folder `biographies` on VUbox. If the folder doesn't contain a biography about the scientist you are working on or the biographic article doesn't provide all the required location information, you should follow the procedures described in section D.4.5.