The Benefits of Forced Experimentation: Striking Evidence from the London Underground Network*

Shaun Larcom† Ferdinand Rauch‡ Tim Willems§

May 16, 2017

Abstract

We present evidence that a significant fraction of commuters on the London underground do not travel on their optimal route. We show that a strike on the underground, which forced many commuters to experiment with new routes, brought lasting changes in behavior. This effect is stronger for commuters who live in areas where the underground map is more distorted, which points to the importance of informational imperfections. Information resulting from the strike improved network-efficiency. Search costs alone are unlikely to explain the suboptimal behavior.

JEL-classification: D83, L91, R41

Key words: Experimentation, Learning, Habit, Optimization, Search

*Many thanks to Transport for London (especially Dale Campbell and Maeve Clements) for providing us with access to the commuter data. We also thank Robert Barro (editor), Maarten Bosker, David Cox, Marco Grazzi, Vernon Henderson, Guy Michaels, Peter Neary, Henry Overman, Jörn-Steffen Pischke, Diego Puga, Collin Raymond, Kevin Roberts, Daniel Sturm, Alastair Young, three anonymous referees, seminar participants at Bern, Bologna, the UK Department for Transport, the London School of Economics, Massachussetts Institute of Technology (Sloan School of Management), Munich, Oxford, the Royal Economic Society Conference, St Andrews, and Toronto for useful comments and discussions. Francesca di Nuzzo provided excellent research assistance. Transport for London does not necessarily agree with the views and conclusions expressed in this publication.

†Department of Land Economy, University of Cambridge. E-mail: stl25@cam.ac.uk.

‡Department of Economics, University of Oxford. E-mail: ferdinand.rauch@economics.ox.ac.uk.

§International Monetary Fund. E-mail: twillems@imf.org.

1 Introduction

Do agents make first-best choices? And do their choices converge to the optimum if the underlying problem repeats itself over time? These questions address fundamental assumptions of economics across all fields, and have been subject to intense debate. We present evidence on these issues by analyzing a unique dataset from the London underground network (also referred to as "the Tube"), which enables us to track individual commuter behavior over time. Our results suggest that a significant fraction of commuters fail to find their optimum commute and that individuals under-experiment in normal times.

Our dataset contains detailed information on commuters using Transport for London ("TfL") services for four weeks. Most commuters use the electronic "Oyster Card" ticketing system to pay for their journeys. Each Oyster Card is associated with a unique anonymized identifier, which allows us to track the movement of individual commuters over time. In the middle of our sample period, the London underground transportation system was subjected to a strike. This caused major disruptions to the network on February 5 and 6, 2014. On these days, some (but not all) Tube stations were closed. As a result, certain commuters were forced to experiment and explore new routes during this period, while others could go to work as usual. We analyze whether such an episode of forced experimentation produces any observable effects that last beyond the duration of the strike. That is: when all stations were open again on February 7, did people who could not use their usual commuting route switch back to their original commuting paths? Or did some of them stick to alternative routes that they may have found during the disruption? By revealed preference, the latter possibility would suggest that they prefer the newly discovered alternative to their old habit, which would indicate that these commuters failed to find their best alternative before the strike.

Commuting itself is an important predictor of life satisfaction. Ahlfeldt *et al.* (2015) for example suggest that a 10 minute commute reduces utility by 14 percentage points. But despite the sizable stakes, we encountered various examples in the media of commuters who changed their route following the strike.[1] This indicates that a sizable share of people might be stuck in sub-optimal habits, even when it comes to activities that are repeated frequently. The number of people who use underground rail networks is also large: there are over 200 subways in

[1] See *e.g.* http://www.bbc.co.uk/news/uk-england-london-26037534, where it is noted that "some commuters have discovered more enjoyable ways of getting to work." That article for example cites a commuter named Andy, who has experimented by taking the Thames Clipper water bus. He commented: "It has been fine, the boat is a great journey. I think I will get the boat back too." Another commuter is quoted as saying that "the walk from Liverpool Street was a refreshing change from the horrors of the Circle Line. I suspect I may permanently switch so I can cut out this, the most stressful part of my journey." After reading a first draft of our paper, a colleague shared the story of his father with us: less than a year before his retirement, renovation works made him discover a new way to his job in Chicago which was superior to his old route that he had been taking for over 40 years.

operation across the world, with the larger systems carrying more than one million passengers per day.[2] Furthermore, Anderson (2014) has shown that even if public transport systems carry only a small fraction of commuters within an urban area, they can have large effects on overall transport system efficiency and the level of congestion.

The results of this study provide evidence on the ability of individuals to find optimal paths in networks. While the rational approach to decision-making has a long history in economics (see the contributions by Rothschild (1974a), Weitzman (1979), Roberts and Weitzman (1981), and Morgan and Manning (1985) to the literature on optimal search), others have remained skeptical of this characterization. Simon (1955) for example argued that agents are "satisficing" rather than maximizing – meaning that they stop their search-for-the-optimum once they have reached a satisfactory utility-level and apply rules-of-thumb from that point onward. Subsequent work by Baumol and Quandt (1964) argues that such behavior may be rational when there are costs associated with decision-making. This anticipates the "costly search literature" pioneered by Rothschild (1974a) and Weitzman (1979).[3] Baumol and Quandt (1964) distinguish between "optimal" and "maximal" solutions: the maximum refers to the exact solution, which would be obtained if there were no search costs, while the optimum takes such costs into account.

Our results also provide evidence on the inclination of individuals to experiment. After all, the new commute was already available pre-strike and could have been found beforehand through voluntary, as opposed to forced, experimentation. Many theoretical papers have pointed out that a certain degree of experimentation is optimal in settings where information is imperfect,[4] but to the best of our knowledge there is no field data-based empirical work analyzing the incidence, as well as the effects, of experimentation in practice.[5] This paper is able to contribute along this dimension: we know when many commuters were experimenting (during the strike), while the Tube-environment provides us with a setting where information is imperfect, making a certain degree of experimentation optimal. The distorted nature of the schematic London Tube map, which many travelers use to navigate, makes it difficult for travelers to minimize journey time (Guo, 2011).[6] Many line-characteristics such as the line's crowdedness and the

[2] London's underground system is the 8th largest in terms of number of stations (270) and 11th largest in terms of passenger numbers (approximately 1.2 billion passenger trips per year).

[3] More recently, Sims' (2003) theory of rational inattention formalizes a similar idea: in his setup, decision makers have to allocate their scarce attention over multiple sources of uncertainty, which leads to deviations from standard maximizing behavior. Also see Matejka and McKay (2015) for an extension of the theory of rational inattention to a discrete-choice setup that characterizes our setting (should I take route A or route B?).

[4] See *e.g.* Rothschild (1974b), Aghion *et al.* (1991) and Bolton and Harris (1999).

[5] Gabaix *et al.* (2006) conducted a laboratory study in which they found that a myopic strategy is a better characterization of actual search behavior than the rational approach set out in Weitzman (1979).

[6] Alan Turing characterized the informational imperfection by describing a friend as "[thinking] of Paris like [...] I would think of a Riemann surface; he only knew the circles of convergence round every Metro station, and couldn't analytically continue from one to another" (Hodges, 2014: 610). Similarly, *The Guardian* of April 27 2015 writes: "When you first move to London it's very common to quickly gain very detailed, even intimate

follow-up journey to the final destination are initially unknown, which adds to the opaqueness of the situation.

Given the presence of informational imperfections, our study may add to the debate on the "Porter-hypothesis". Porter (1991) argued that – when information is imperfect – exogenously-imposed constraints may help agents to get closer to their optimum by triggering a period of experimentation, innovation and re-optimization. Porter originally phrased his hypothesis in the context of environmental regulation, but the underlying idea is more general and also applies to the London underground setup considered in this paper.[7] The Porter-hypothesis imposes a degree of irrationality on the part of decision-makers. After all, why would it take an exogenously-imposed constraint to make agents realize that they were not optimizing beforehand? Why wouldn't they experiment voluntarily? As a result, Porter's hypothesis has been dismissed by many scholars as being unrealistic – initially mostly on anecdotal grounds (see *e.g.* Palmer, Oates and Portney [1995] and Schmalensee [1993]). Subsequently, many studies have tried to test the theory empirically but, as noted by Porter and Van der Linde (1995) and Ambec *et al.* (2013), data limitations make it hard to put Porter's hypothesis to a proper test in practice. The fact that measureable progress often takes time to occur makes it for example difficult to keep "all else equal", while it is also not clear how "an improvement" could be defined in the first place. As a result of these complications, the literature has not settled upon a consensus with respect to this issue (see *e.g.* Gray [1987], Jaffe and Palmer [1997], Berman and Bui [2001], and Copeland and Taylor [2004]). By analyzing the behavior of commuters who were faced with a short-lived, temporary constraint on the London underground network, the present study overcomes some of these problems.

In addition, our study is informative on the existence, strength and persistence of habits. As noted by Wood and Neal (2009), research on habits, or a preference for the familiar, is important since about 45% of people's behavior is repeated on a daily basis. Commuter behavior is an example of this. Along these lines, Goodwin (1977: 95) has argued that "the traveler does not carefully and deliberately calculate anew each morning whether to go to work by car or bus. Such deliberation is likely to occur only occasionally, probably in response to some large change

knowledge of two or three locales, but not know how they are connected geographically. It's not until there's a Tube strike and you have to cycle or take the bus [...] that you suddenly realize that places you thought were separated by several sets of escalators and two Tube lines are only 15 minutes walk apart. It was only last week that one of us realized that Goodge Street is a short walk from Euston Station."

[7] Porter stated that tighter environmental standards "do not inevitably hinder competitive advantage against foreign rivals; indeed, they often enhance it. Tough standards trigger innovation and upgrading". Similarly, Porter and Van der Linde (1995: 98) claim that environmental regulations can "trigger innovation (...) that may partially or more than fully offset the costs of complying with them". This idea goes back to the notion of "induced innovation", developed in Hicks (1932), and has also been taken beyond Porter's original application to environmental regulation (see *e.g.* Aghion, Dewatripont and Rey [1997] for a paper that analyzes related issues in a more general setup).

in the situation".

Finally, our paper provides evidence on the effects of a public transport strike. Although there are some earlier studies analyzing disruptions in transportation networks (see Van Exel and Rietveld (2001) for an overview of this sparse literature), they tend to rely on survey data – leading to small sample sizes and preventing a clean comparative analysis of travel patterns before and after the disruption (Zhu and Levinson, 2011: 19). As we will explain in greater detail in Section 4, the present study has the entire population of actual travel movements on the London underground at its disposal, which brings advantages over earlier contributions.

The remainder of this paper is structured as follows. We start by providing background information on the London underground network in Section 2. Section 3 describes the Tube strike that took place in February 2014. We discuss our dataset in Section 4. To motivate certain choices in our empirical exercise, we provide descriptive statistics on commuters in Section 5. Section 6 describes our estimation method. We present our analysis of the effects of the strike in Section 7. In Section 8 we interpret our results using a model of rational search. Section 9 concludes.

2 The London underground network

Over the sample period considered in this paper, January 19 to February 15, 2014, the London underground network consisted of 11 different lines, connecting 270 different stations. It is operated by London Underground Limited, which is fully owned by Transport for London, the corporation that runs most of London's public transport services. It covers 402 kilometers of track. The London underground serves up to 4 million passenger journeys per day and is a popular mode of transportation for many people living and working in London.

Crucially for our paper, users of the London underground face imperfect information on several relevant features of the available alternative routes in getting from A to B. An important source for this imperfection is the London Tube map – a major aid to travelers in finding their way through the network. It is a schematic transit map, showing only *relative* positions of Tube and train stations along lines. Consequently, the map is geographically distorted and gives users false impressions when it comes to actual distances between two points – especially when comparing points along different Tube/train lines.[8] The distorted nature of the map gives rise to further problems of similar nature when traveling from the exit station to the final

[8] Guo (2011) calculates that for the London underground map, the correlation between actual and "mapped" distances is only 0.22. He also gives several examples of actual distortions. A famous case is that of Covent Garden and Leicester Square: both stations are only 260 meters apart, but the 20 second Tube ride remains in high demand.

destination, which is likely to lie somewhere in between the various lines, where the map is not well-defined.

Next to commuting time, travelers are initially also uncertain on many characteristics of the various available alternatives. How crowded is a particular line at the preferred time of travel? Is the route from the exit station to the final destination convenient? There could for example be a supermarket along the way, or a place that serves good breakfast.

An important way in which these various uncertainties can be reduced, is by actually trying the available alternatives, that is through experimentation. And because of the strike that we are about to describe in the next section, many travelers were forced to do exactly that in the first week of February 2014.

3 The strike

On January 10, 2014, the Rail Maritime Transport union, the largest trade union in the British transport sector, announced a 48-hour strike of London Tube workers. The strike was scheduled to begin on Tuesday evening (21:00h) 4 February. It was called for in response to the announcement of a plan by Transport for London to close ticket offices and to introduce non-compulsory redundancies for part of its workforce.

The decision to participate in the strike remained with individual workers. In the past, it has therefore sometimes been the case that unions called for a walkout, but workers did not act accordingly. For example, in December 2005, the union called for action on New Year's Eve but, according to an official bulletin, the "strike has had little impact on London Underground's services (...) The majority of station staff have ignored the call for industrial action and are working normally."[9]

However, more workers participated in the February 2014 strike. Due to the resulting non-availability of staff members, 171 (out of 270) Tube stations remained closed for at least part of the duration of the strike (see Figure I for a visualization). Which stations were closed was decided in a spontaneous process heavily influenced by the availability of staff and was not known until the morning of the strike. Many of the factors were outside the control of Transport for London, such as the degree of unionization along various lines/stations and the willingness of individual employees to participate in the strike. The underlying process does not seem to have been strategic. As a result, treatment and control stations are similar along observable characteristics. Most importantly, the number of neighboring stations within walking distance

[9]See tfl.gov.uk/info-for/media/press-releases/2005/december/london-underground-service-update –2000hrs.

(1km, 2km, and 5km) is not statistically different between treatment and control stations.[10] As our main outcome is the probability to switch commuting route, this is the key dimension along which our exercise requires similarity between closed and open stations.

— **Figure I about here.** Figure caption: *Impact of the February 2014 Tube strike. Circles represent stations on a standard Tube map (includes Overground and DLR) with GPS coordinates used to locate position. Crosses represent stations that were fully closed during the strike period.*

There are a number of stations on the network that serve multiple lines and were only partially closed during the strike, with one or more lines still operating on them. In our econometric exercise we code these stations as closed, even though some commuters might have been able to use them. During the two strike days, many lines had fewer trains running, while there were no services on the Bakerloo line, the Circle line, and the Waterloo & City line. For a large part, these closed lines run in parallel and overlap with lines that remained open. As of Friday morning February 7, all stations were open again with services running as usual.

The February 2014 strike has several desirable features that make it particularly suited for studying the question at hand. First, it was the first major disruption in over three years. As a result the sample is likely to contain many individuals who hadn't been subjected to forced experimentation on such a scale before.[11] Second, the strike was not complete. About 37 percent of all stations remained open, which enabled travelers to experiment within the Tube network. This would not be possible if all stations were closed. Third, the partial nature of the strike furthermore leaves us with a sizable share of both "treated" and "non-treated" commuters, which will help our analysis below. Fourth, the first full strike day (February 5) was rainy. According to weatheronline.co.uk there were 7mm of rain in London during the morning, which is likely to have discouraged travelers from experimenting by bike or foot, in which case they would no longer show up in our data and we no longer know whether they went to work or worked from home.[12] Finally, with a duration of 48 hours, the strike was relatively short-lived. Thus any changes in behavior are likely to be driven by optimality-considerations – not by the formation of habits during the strike, which we believe would take longer to form.

[10]The p-values of tests of similarity are 0.76 (1km), 0.54 (2km) and 0.12 (5km).

[11]London attracts about 350 thousand new inhabitants per year. This implies that about 1.2 million Londoners at the time of the most recent disruption had not been living there during the previous strike in 2010. Others would have changed jobs or houses in the meantime – resetting any previous commute-specific information they acquired.

[12]Also see the advance warnings, for example "Weather hits trains as London Tube strike begins" in *The Guardian* of February 4, 2014.

4 Data

Data on commuting behavior are particularly well-suited to analyze decision-making and learning by individuals. First, commuters are faced with a very similar problem, to get from home to work and back, on a workdaily frequency. Our dataset shows us, in great detail, how they answer this problem at the exact same frequency. Second, many aspects of the problem, such as journey time, are quantifiable and observable to us, which allows for a rich analysis. Third, the solution to the problem is in many cases not trivial: there are for example no less than 13 reasonable ways to travel from King's Cross to Waterloo. Finally, the fact that only part of the network closed down during our sample period leaves us with treated and non-treated individuals – which greatly facilitates a comparative analysis.

The dataset that we use was provided to us by Transport for London. It contains all individual travel movements on the London public transport system from Sunday January 19 to Saturday February 15, 2014. For all modes of public transportation other than bus (Tube, train, tram, boat, and DLR – the Docklands Light Railway, a train transit system that is part of TfL), the dataset provides us with the station of entry for a particular journey, the station of exit, as well as the times of check-in and check-out.[13] The February 2014-strike affected only the Tube network. All boat, bus, train, tram, and DLR stations remained in operation. Thus the focus of our study is on journeys that involve the underground.

Over our sample period, payments for individual journeys could be settled in two ways: either by purchasing a ticket that is valid for a certain time period and area, or by using a re-chargeable plastic card called an "Oyster Card". In the period we study, Oyster Cards were used in about 80 percent of journeys. Each Oyster Card has a unique anonymized number, which allows us to track individual travel behavior over our sample period. As we want to observe how repeat-behavior changes after a disruption, we analyze Oyster Card-using commuters who face the same problem of getting from A to B and back every weekday.

We identify Tube commuters as individuals who use London's Tube network during every of the 12 pre-strike working-days in our sample between 7am and 10am. The presence-requirement leaves us with a panel of Tube-users, while the time-requirement implies that we only look at the morning rush hour, which runs from about 7am to 10am, see Figure III below. We want to focus on commuters that are stuck in a daily routine, which is more likely in the morning than in the evening, when Londoners may pursue other activities before heading home.

We create two main datasets, which we call "unconditional" and "balanced". In the unconditional dataset we use no information from the post-strike period to identify our commuters.

[13] TfL does not record exit stations for buses as bus passengers pay a flat fee.

We only require individuals to be present in our dataset on every morning during the pre-strike period. The advantage of this dataset is that there is no selection of data based on outcomes. In the balanced sample, on the other hand, we require individuals to be present on every morning in both the pre- *and post-strike* period, as well as on at least one strike day.[14] The advantage of this balanced panel is that it enables us to observe behavior of all commuters included in our dataset before, during and after the strike, as by construction, no one abandons the TfL-system. The disadvantage of this approach is that we require the commuter to remain within TfL-services, which is a selection on an outcome. As we will show in Section 7.1, treated and untreated individuals are about equally likely to opt-out of the TfL-system post-strike. This suggests that the selection procedure underlying the balanced sample is not problematic. We therefore proceed by using the balanced sample as our main one – taking full advantage of the fact that this sample allows for a detailed comparison of behavior before and after the strike.

We infer the "usual" entry and exit stations of travelers by setting it equal to the station which they use most frequently during the pre-strike period (the "modal station"). A small minority of about 700 individuals have multiple modes on either or both ends that are used with equal frequency. These create complications in our analysis, which is focused on identifying "deviations from the mode" – assuming the latter is unique. We drop them as well. Cutting the data in this way leaves us with an unconditional sample of 60,747 Oyster Card IDs, and a balanced panel of 17,343 Oyster Card IDs. For each of these IDs we have 20 working days of observations.

Throughout this paper, we employ a rather strict definition of the concept of a "commuter" as we require them to behave in a very consistent manner. Consequently, we exclude individuals who would be called commuters by other definitions. We also miss individuals who use multiple Oyster Cards, as well as those who were absent from London's public transport system for one non-strike weekday or more over our sample period. Given the size of our data, these misses still leave sizable sample sizes. Moreover, if anything, this strict selection procedure implies that the mode-change probabilities reported below are a lower bound, as we have selected those individuals who adhered to some routine during the pre-strike period.

[14]The latter requirement serves to ensure that we analyze the behavior of individuals who were actually present on the underground during the disruptive phase (instead of working from home) – thereby making sure that they have had a chance to explore alternative routes during this period.

5 Descriptive statistics

We start by providing some descriptive statistics based upon the data on the whole population of London commuters. Next to that, these statistics are used to motivate certain choices that we make in the econometric exercise that is to follow. Our data are informative on the dominant public transport commuting patterns within the Greater London area. Figure II displays stations of first entry in the morning and evening for one of the days in our sample, January 31, 2014. Circle-sizes correspond to relative station use. The morning commute is characterized by a dispersed start, often from residential areas in the outskirts of London or the large commuter railway stations on London's periphery. The evening commute, on the other hand, is much more concentrated – starting from well-known business districts like Canary Wharf and the City.

— **Figure II about here.** Figure caption: *Stations of first entry in the morning and evening of January 31, 2014.*

Second, due to the absence of other significant events during our sample period all non-strike working days were approximately equally busy:[15] the busiest day is Friday January 24, 2014 (with 19,301,730 data entries and 3,652,851 unique travel IDs) while the quietest non-strike day of our sample is Wednesday February 12, 2014 (with 18,259,114 data entries and 3,496,720 unique travel IDs). Within each day, activity followed a standard pattern, as displayed in Figure III for January 31st. The figure shows that the morning commute runs from about 7am to 10am, which motivates our earlier choice along these lines.

— **Figure III about here.** Figure caption: *Travel pattern of January 31, 2014. The horizontal axis represents time (hours) and the vertical axis represents travel volume (density).*

— **Figure IV about here.** Figure caption: *Summary statistics. The top-left panel shows the fraction of commuters of the balanced panel who enter at their modal station. The top-right panel shows the fraction of commuters of the balanced panel who exit at their modal station. The bottom-left panel shows the duration of the average journey time (minutes). The bottom-right panel shows the degree of dispersion in journey times (standard deviation across commuters). The horizontal axes represent days and the two strike days are located in between the vertical lines.*

Figure IV shows the evolution of some key variables of interest for all weekdays in our sample

[15] Around January 28, which corresponds to day 9 in Figure IV, there was heavy rain and flooding in southern England, which might have influenced travel decisions of some commuters. During our sample period, no other announcements of travel disruptions on the London Underground other than the Tube strike could be found.

period. The top-left panel shows the fraction of commuters by the balanced panel who enter at their modal station, while the top-right panel shows the same at the exit-margin. The two strike days can be found in between the vertical lines. As the two panels show, fewer commuters were able to use their modal station during the strike – which implies that a substantial number of individuals were forced to explore alternative routes. Moreover, the post-strike data also suggest that the strike brought about some lasting changes in behavior, as the fraction of commuters that make use of their modal station can be seen to drop after the strike.[16] The lower two panels of Figure IV provide information on journey times: as the bottom-left panel shows, the duration of the average journey on London's public transport system went up during the strike (by about 6%), while the bottom-right panel shows that this increase in average duration was also accompanied by an increase in dispersion.

6 Estimation Strategy

As set out before, the partial nature of the February 2014 strike conveniently leaves us with treated and non-treated commuters. This enables a difference-in-differences exercise, which is the approach that we take in this paper. We typically estimate regression equations that are of the following form:

$$d_{it}^{\text{mode}} = \alpha_i + \beta \cdot d_t^{\text{post}} + \gamma \left(d_t^{\text{post}} \cdot d_i^{\text{treat}} \right) + \epsilon_{it}. \tag{1}$$

Here, d_{it}^{mode} is a dummy-variable that takes the value 1 if individual i makes his "modal journey" (travels from his modal station of entry to his modal station of exit) on day t, d_t^{post} is a dummy-variable that takes the value 1 in the post-strike period, while d_i^{treat} is a dummy-variable that takes the value 1 if individual i was part of the treatment group, defined in different ways as described below. ϵ_{it} in equation (1) is the error-term, β measures time effects, while γ captures the treatment effect. Data from the two strike days are not included in the estimations, and used only to identify the treatment group. We estimate equation (1) using OLS and apply robust standard errors. When treatment is defined at the station level we cluster standard errors at this level. We include individual fixed-effects as captured by α_i in equation (1). The reason for the inclusion of individual fixed effects is threefold. First, fixed effects control for unobserved demographic factors, such as age, that may affect an individual's inclination to experiment. Second, they control for station or area characteristics relating to the modal stations of an individual, which may influence the propensity to switch. Third, fixed effects correct for the fact that different individuals use their modal station with different intensity.

[16] Establishing this formally is the objective of the remainder of this paper.

Since the presence of fixed effects in combination with interaction terms renders the estimation of a discrete choice model econometrically challenging, we employ a linear probability model instead.

As we will clarify in the remainder of this section, identifying the treatment group from our dataset is not straightforward. To ensure robustness, we will therefore show results for three different definitions of treated commuters – where all measures have their specific advantages and disadvantages.

Our first measure of treatment defines treated individuals as all those who deviate from their pre-strike modal journey during the strike, in terms of stations used. This includes individuals who were forced to explore a new route due to closure of an entry, exit, or connecting station, but will also encompass those who deviated from their pre-strike mode for non-strike related reasons. The second measure of treatment takes a more direct approach: in this exercise, we take individuals to be treated if their pre-strike modal station, either entry or exit, closed down during the strike. These individuals were not able to travel to or from their modal station during the strike and hence forced to explore alternatives. This measure does not allow for selection into treatment, as it seems reasonable to assume that the closure of stations is random with respect to individual characteristics. It suffers from the fact that it is likely to pool a significant number of treated individuals with the non-treated group. The reason is that many individuals in our dataset travel from station A to station B via at least one connecting station C. Closure of the latter would force this individual to explore alternatives, but we don't observe connecting stations in our data, only stations of entry and exit. Consequently, our second measure of treatment is likely to lead to type II errors and underestimate the true effect. Our third measure of treatment is based on travel time: here we take individuals to be treated if their travel times during strike days were sufficiently different (i.e.: longer or shorter) from their travel times during the pre-strike period. This method identifies those commuters who had an unusual experience during the strike as measured by time. It does not rely upon our definition of closed stations (as pointed out in Section 3, some stations were only partially closed), while it also side-steps our concept of "deviations from the modal commute" as used in the other treatment definitions. This measure of treatment is however prone to errors of both the first and the second kind (i.e.: there will be both "false positives" and "false negatives"). After all, if an individual had a different journey time on strike days, that does not necessarily imply that he was actually exploring an alternative route. It could simply be the case that his modal route took much longer due to congestion on the network, or due to the reduction in the number of trains running. Similarly, it is also possible that a commuter explored a different route, but that this did not lead to a markedly different travel time. Each treatment measure is summarized in Box 1 below.

Box 1: Treatment measures.

Treatment name	Treatment definition
1. Different station	Commuters who deviated from their pre-strike modal station during the strike.
2. Station strike	A commuter's pre-strike entry or exit modal station is closed during the strike. "Entry on strike" (2a) refers to the case in which a commuter's pre-strike modal station of entry was closed during the strike, while "exit on strike" (2b) refers to the pre-strike modal station of exit being closed during the strike.
3. Time factors	Travel times of commuters during strike days were different (longer or shorter) from their travel times during the pre-strike period; by a factor of 1.2 [3a],1.5 [3b], or 2 [3c].

For all treatment measures we verify that the treatment and the control group are similar for observables, such as average journey duration, and the number of stations near modal entry and exit station. Figure V displays trends for treatment and control groups for our outcome variables, taken from the balanced sample. The graph shows the fraction of commuters who travel using both their modal entry and exit station. Visually, the pre-strike levels and trends look similar for both groups in all cases, with the exception of the "different station" group, where levels in the pre-strike period differ somewhat.[17] This suggests that commuters who switch during the strike, are more likely to deviate from their mode before the strike as well. This points to some selection into treatment for this variable. This may be offset by including the treatment dummy as a separate variable in the estimation equation. It is furthermore not clear whether any residual that is left after accounting for the treatment dummy, biases our estimates in the up- or downward direction (if at all). On the one hand, it could point to easier substitution for the treatment group in which case we would expect them to have a greater propensity to switch if treated. On the other hand, it may also be the case that the treatment group experimented more pre-strike, uses better routes to start with as a result, and thus has a smaller propensity to change behavior post-treatment. For all the other treatment definitions, the pre strike levels and trends look very similar.

— **Figure V about here.** Figure caption: *Time trends for treatment and control group for each measure of treatment (defined in Box 1). For each of the panels the horizontal axes represent days and the vertical axes represent the fraction of commuters of the balanced panel*

[17]The treatment definition employed in this case implies that, by construction, the fraction of commuters in the control group traveling their pre-strike modal commute jumps to 1 on both strike days.

who enter at their modal station.

7 Findings

In this section we present our main estimation results. Section 7.1 describes the main regression results, Section 7.2 provides some robustness checks for these results, Section 7.3 analyzes the effects on travel time, while Section 7.4 investigates the mechanism underlying the main effect.

7.1 Core results

As set out in Section 6, we rely upon difference-in-differences estimations to ask whether treated commuters were more likely to deviate from their pre-strike modal journey in the post-strike period, relative to their non-treated peers. The answer to this question can be found by looking at the sign of our estimate of the treatment effect γ in regression equation (1).

We start by considering whether the strike affected people's decision to abandon the public transport system as a whole. Here we introduce an alternative left hand side varialbe, s_{it}, which indicates presence in the TfL-services, excluding buses. It equals "1" for commuters who use the Tube, train, boat or the DLR, while it equals "0" for people who use all other modes of transportation or stay at home.

Table I: OLS-DiD results for commuters remaining in the Tube system (unconditional sample).

	(1: different station)	(2: station strike)	(3a: factor 1.2)	(3b: factor 1.5)	(3c: factor 2)
β	-0.163***	-0.162***	-0.147***	-0.146***	-0.147***
	(0.001)	(0.0007)	(0.0009)	(0.0007)	(0.0006)
γ	-0.0021	0.0017	-0.021***	-0.032***	-0.041***
	(0.0012)	(0.0011)	(0.0011)	(0.0011)	(0.0012)
obs	1,093,446	1,093,446	1,093,446	1,093,446	1,093,446

Notes. Table I reports OLS estimates of equation (1) where the dependent variable indicates presence of commuter i on day t on TfL-services, excluding buses. The columns present the results for each measure of treatment (defined in Box 1). β is the coefficient of the post-strike dummy (d_t^{post}) and γ is the coefficient of the interaction term between the post-strike and treatment group dummies ($d_t^{post} \cdot d_i^{treat}$). * denotes significance at the 10% level, ** implies significance at the 5% level, *** indicates significance at the 1% level. Individual fixed effects are applied in each estimation and robust standard errors are recorded in the parentheses.

Table I uses the "unconditional" sample. In column 1 the treatment group consists of those people who deviated from their usual route on either strike day. The coefficient on the post-strike dummy β is large and significantly negative, which we would expect given that the modal station is identified using information from the pre-strike period only. Other columns of Table I confirm this. The coefficient on the interaction of post-strike and treatment dummies γ is not statistically significantly different from zero. This suggests that for this measure of treatment, people do not have a greater propensity to opt out of the system if they had to change their commute during the strike. The second column identifies the treatment group as those commuters whose modal commute started from or ended at a station which participated in the strike. Again the interaction-coefficient is not statistically different from zero. In the three remaining columns we measure the treatment by commuters who had a substantially longer or shorter journey on strike day. Here, we find significant negative treatment coefficients. This suggests that some commuters who had bad experiences during the strike indeed decided to abandon TfL-services. The propensity to abandon is increasing in the intensity of journey time deviation on strike day (the estimate of the absolute value of γ is larger in column 3c than it is in column 3b, which is in turn larger than column 3a). For the time-based definitions of treatment, the fact that our estimate of γ is negative suggests that any treatment effects on switching behavior reported below are likely to form a lower bound, since they are based upon a sample which excludes the group of commuters that switched radically by ceasing to use the TfL-system post-strike. Our main measures, in the first two columns, show no effect. This suggests that selection into other means of transportation is not a great concern for these measures, which is why we proceed with the balanced sample for the rest of the analysis.

Table II reports our main estimates of interest – employing the same treatment-definitions as in Table I. In all specifications, the interaction coefficient γ (measuring the difference-in-differences) is consistently estimated to be significantly negative. This suggests that those who were forced to explore alternatives during the strike, were less likely to use their pre-strike modal commute after the restriction was lifted.[18] By a revealed preference-type argument, this suggests that a significant fraction of commuters had failed to find their optimal journey before the strike. Post-strike, all routes were available again, including the pre-strike modal one, so a failure to pick the latter option suggests that the commuter has found a better alternative during the disruption. Our results are unlikely to be driven by the formation of new habits during the two strike days. Not only do they typically take much longer to be established (Wood and Neal, 2009), but the observed behavior of commuters is also inconsistent with this hypothesis: after the strike, many of them continue to explore alternative routes (leading to a prolonged experimental phase) after which they eventually settle on a new modal choice.[19]

[18] In the specification where we identify the treatment group via the time factor (Columns (3a), (3b) and (3c)), this effect is consistently stronger for those travelers who experienced a shorter deviation from their usual commuting time during the strike, which is intuitive: the alternative route is likely to look less attractive to a commuter if (s)he experienced a long delay during the strike.

[19] To give a random example: one commuter in our dataset consistently traveled from station S to station C in the pre-strike period (for privacy-reasons, we are not allowed to give full station names). During the strike, (s)he

Table II: OLS-DiD results for commuters using modal station within the Tube system.

	(1:different station)	(2: station strike)	(3a: factor 1.2)	(3b: factor 1.5)	(3c: factor 2)
	Mode Station	Mode Station	Mode Station	Mode Station	Mode Station
β	-0.0108***	-0.0466***	-0.0402***	-0.0464***	-0.0504***
	(0.00186)	(0.00185)	(0.00175)	(0.00140)	(0.00128)
γ	-0.0569***	-0.00860***	-0.0205***	-0.0201***	-0.0113**
	(0.00242)	(0.00248)	(0.00245)	(0.00293)	(0.00451)
obs	312,156	312,156	312,156	312,156	312,156

Notes. Table II reports OLS estimates of equation (1) where the dependent variable indicates whether commuter i on day t traveled using their modal station (entry or exit). The columns present the results for each measure of treatment (defined in Box 1). β is the coefficient of the post-strike dummy (d_t^{post}) and γ is the coefficient of the interaction term between the post-strike and treatment group dummies ($d_t^{post} \cdot d_i^{treat}$). * denotes significance at the 10% level, ** implies significance at the 5% level, *** indicates significance at the 1% level. The share of treated commuters is 0.71 for Column (1), 0.55 (2), 0.54 (3), 0.24 (4), 8.39 (5). Individual fixed effects are applied in each estimation and robust standard errors are recorded in the parentheses.

Looking at the magnitudes of estimates across the various tables is informative as well. A negative estimate for β is to be expected due to natural changes in workplace or home and a definition of the modal station that is based on the pre-strike period only. However, a large negative estimate for β could also suggest that our treatment-definitions err by including treated individuals in the non-treated group. As anticipated in Section 6, we expect our second measure of treatment to be particularly prone to this statistical error of the second kind – and indeed, the absolute value of the estimate of γ is lowest in this specification, while that of β is among the highest. In column 1 on the other hand, the estimate of β is closer to zero, which makes sense from a theoretical point of view, as a result of which this table contains our preferred estimates for the treatment effect γ. The key magnitude is the estimate of γ, which equals -0.057 for this definition of the treatment group. It suggests that the probability to switch on a given journey is five percentage points greater in the post-strike period.

To look at the findings of Table II in a different way, we can increase the level of detail and estimate a specification that distinguishes between entry and exit (so the LHS-variable in that regression is either $d_{it}^{\text{entry\_mode}}$ or $d_{it}^{\text{exit\_mode}}$). Results of this exercise, recorded in Table III, indicate that the treatment effect γ tends to be bigger at the exit-end. This is intuitive since the exit-end of the morning commute typically lies in the city center (Figure II) where station-

experiments with entering at E – using the DLR to travel to C. In the post-strike period, (s)he first alternates between both options (seemingly comparing them) after which (s)he settles for the newly-found DLR-based route. There are also more determined examples: another commuter consistently travels from R to J on every morning before the strike. Both stations however closed down during the strike, in response to which (s)he switched to traveling from N to W on the first strike day. Subsequently, (s)he sticks with this new alternative (which has a shorter duration and a lower variance) for the remainder of our sample period. As direct evidence for experimentation, we also compute a left hand side dummy variable that identifies commuters who enter and exit at the same station as on the previous day. For all five treatment definitions, we get negative and significant treatment effects, with a magnitude between -0.01 and -0.4. This suggests that a sizeable fraction of those commuters who had to reconsider their journeys during the strike, experiment in the days that followed it.

density, and hence the possibility to substitute stations, is higher. Also note that Table III contains only two estimates that are not significant at any regular level of significance (whereas all other estimates are significant at the 1% level). However, they show up at exactly those places where this is plausible, namely when we look at what closure of a modal entry station does to the choice of station of exit, and vice versa.

The coefficients in Tables I and II are fairly straightforward in their interpretation due to the probabilistic nature of our exercise. Some complication arises as commuters in our pre-strike sample only make their modal journey for about 84% of the time: an estimate for γ of -0.03 therefore implies that treated individuals will make their pre-strike modal commute with a probability that is 3 percentage points lower compared to their non-treated peers. This does however not imply that 3% is also the fraction of switchers in our sample.

Table IV, on the other hand, does produce information on the fraction of switchers – as such a number is arguably easier to interpret. This table is constructed by first identifying those commuters who made the exact same morning commute (as far as stations of entry and exit are concerned) during all 12 working days of our pre-strike sample. Hence, all these individuals (who we refer to as "pre-strike habituals") are selected so that they make their modal commute with probability 1 in the pre-strike period. We subsequently ask: how many percentage points higher is the fraction of "post-strike switchers"[20] in the treatment group relative to the fraction of switchers among non-treated commuters?[21]

[20] "Switchers" are defined as those individuals who made a different commute than their pre-strike modal journey on the last working day of our sample (Friday February 14). This exercise therefore assumes that the experimentation phase, triggered by the strike-induced forced episode of experimentation, was over by this time. Requiring them to deviate for more than one day, yields very similar results.

[21] Again we look at deviation probabilities relative to a non-treated control group since this exercise is obviously prone to "regression to the mean". Given that the habituals were using their modal station with probability 1 in the pre-strike period, they can only make (weakly) less use of it post-strike. The control group of non-treated commuters allows us to correct for mean reversion.

Table III: Estimates of γ when distinguishing between entry and exit margin.

Treatment definition	Entry Mode	Exit Mode
1. different station	-0.0267***	-0.0470***
	(0.00164)	(0.00224)
2. station strike	-0.00480***	-0.00859***
	(0.00170)	(0.00229)
2a. entry on strike	-0.00697***	0.000569
	(0.00190)	(0.00247)
2b. exit on strike	-0.00146	-0.00748***
	(0.00173)	(0.00230)
3a. time factor(1.2)	-0.0141***	-0.0154***
	(0.00168)	(0.00227)
3b. time factor(1.5)	-0.0111***	-0.0175***
	(0.00207)	(0.00270)
3c. time factor(2)	-0.00766***	-0.0113***
	(0.00329)	(0.00411)
obs	312,156	312,156

Notes. Table III reports OLS estimates of γ from equation (1). γ is the coefficient of the interaction term between the post-strike and treatment group dummies ($d_t^{post} \cdot d_i^{treat}$). The dependent variable in column 1 indicates whether commuter i on day t traveled using their modal station of entry, and column 2 station of exit. Each of the rows report estimates for different measures of treatment (defined in Box 1). This table is based upon 14 regressions of the same form. For space-constraints, we only report our estimates of γ. Estimates of other coefficients are available upon request * denotes significance at the 10% level, ** implies significance at the 5% level, *** indicates significance at the 1% level. Individual fixed effects are applied, robust standard errors are recorded in parentheses.

Table IV: Fraction of switchers among pre-strike habituals.

Treatment definition	Share treated	Share of switchers
1. different station	0.22	0.054
2. station strike	0.42	0.026
3a. time factor(1.2)	0.44	0.012
3b. time factor(1.5)	0.25	0.019
3c. time factor(2)	0.09	0.028
obs	6,946	6,946

Notes. Table IV reports the proportion of individuals that made a different commute from their pre-strike modal journey on the last working day of the sample period (Friday February 14). Each of the rows report estimates for different measures of treatment (defined in Box 1). These estimates are obtained from regressions identical to Table II. However, to obtain an interpretable estimate of the fraction of switchers, the sample of commuters is limited to those who made the exact same morning commute (in terms of entry and exit stations) during each of the 12 working days of the pre-strike sample period.

As can be seen from Table IV, our data suggest that depending on whom we consider to be treated, the fraction of post-strike switchers is 1.2 to 5.4 percentage points higher in the treatment group. Since results for our last two measures of treatment ("station strike" and the method using the "time factor") are again likely to be biased by type II errors, we believe that the true number lies closer to 5.4 percentage points, the number we obtain when defining the treatment group as those who deviated from their modal journey during the strike. This is a strong result given that the individuals underlying this exercise all seemed to be stuck in a very regular habit before the strike, as they were selected exactly because they were making the same commute on every single of the twelve mornings in the pre-strike sample. The selection method could furthermore imply that these commuters have only few viable alternatives available, which also biases the results against switching. Moreover, exploring a new route during a Tube strike is typically not a pleasant experience, due to the associated chaos and crowdedness, while there were also fewer trains running during the February 2014-strike – causing further delays. Consequently, it is likely that numbers would have been even larger after considering voluntary experimentation under non-strike conditions. In line with our earlier findings, this again provides evidence that a substantial proportion of commuters had failed to find their maximum before the Tube strike of February 2014.[22]

[22]We are not claiming that these commuters have found their global maximum post-strike: all we are saying is that they have found something better than their pre-strike mode, but it is very well possible that further improvements are still possible.

7.2 Robustness

Our estimates of γ are strongly statistically significant across different specifications. Section 7.1 illustrated this for different definitions of the treatment group, while this section shows robustness with respect to alternative regression specifications. This can for example be seen from Table V.

Table V: Estimates of γ across specifications.

Treatment definition	(1: BDM)	(2: SL)
	Mode Station	Mode Station
1. different station	-0.0569***	-0.0414***
	(0.00311)	(0.00540)
2. station strike	-0.00860***	-0.0208***
	(0.00333)	(0.00608)
3a. time factor(1.2)	-0.0205***	-0.0116**
	(0.00331)	(0.00568)
3b. time factor(1.5)	-0.0201***	-0.0223***
	(0.00406)	(0.00664)
3c. time factor(2)	-0.0113*	-0.0337***
	(0.00615)	(0.0105)
obs	34,684	47,052

Notes. Table V reports estimates of γ from equation (1) for two different specifications. γ is the coefficient of the interaction term between the post-strike and treatment group dummies ($d_t^{post} \cdot d_i^{treat}$). Column 1 reports estimates of γ where the data are aggregated into two observations for each individual: one observation pre-strike and one post-strike. Mode refers to the mean number of modal journeys before and after the strike. Column 2 reports estimates of γ where the sample is restricted to those individuals who enter and exit on the same line. This table is based upon 10 underlying regressions. Due to space-constraints, we only report our estimates of γ. Estimates of other coefficients are available upon request. Each of the rows report estimates for different measures of treatment (defined in Box 1). * denotes significance at the 10% level, ** implies significance at the 5% level, *** indicates significance at the 1% level. Individual fixed effects are applied, robust standard errors are recorded in parentheses.

A well-known criticism of OLS-DiD panel data regressions, is that autocorrelation for individuals over time biases estimated standard errors towards zero (Bertrand, Duflo and Mullainathan, 2004). In column 1, we therefore report results generated by their most conservative robustness check – one where the data are aggregated to two observations for each individual: one observation pre-strike and one post-strike. Here we aggregate our LHS variable by computing the mean number of modal journeys before and after the strike. Given that all variables used in our specification are binary indicator variables, coefficients remain numerically identical in this exercise. They still are highly statistically significant.

Column 2 shows our baseline estimates of γ when we restrict our sample to those individuals who enter and exit on the same line ("SL"). As set out before, identifying the treatment group is somewhat challenging in the full sample as many individuals make use of connecting stations during their commute. Closure of a connecting station implies that such an individual was treated during the strike, even if his entry and exit station remained open. We do not observe data on connections. This concern plays no role when we limit ourselves to those commuters who enter and exit on the same line, as they are unlikely to travel via a connecting station. Due to the "same line" restriction we are left with fewer observations. Our main results persist. The estimates are less significant, which is no surprise given the smaller sample size. They are typically smaller in magnitude, likely due to less scope for substitution for one line commuters. The estimated coefficients are numerically closer to each other, possibly due to the measurement problems that lead us to run this specification.

In a further robustness check, we interact each day in the post-strike period with each of the treatments separately, to study the timing of the treatment effect. We find coefficients of similar magnitudes for all these days, and no clear time trend in either of the specifications.

7.3 Effects on travel time

A follow-up question to ask at this stage concerns the effect of the strike on commuting times. We do not observe the duration of the entire commute, since commuter behavior is not observed outside of TfL-services. We can however analyze the amount of time they spent on London's public transport network. As we do not know whether the time commuters spend traveling to and from the stations correlates with the observed journey times in the TfL-systems, these results need to be interpreted with caution. However, since time spent on the London underground during rush hour is well-known to be particularly unpleasant, also compared to other modes of commuting, minimizing this time is likely to receive a significant weight in the objective functions of most commuters. Moreover, these numbers should at the very least be of interest to the providers of public transport services, as everything equal they have an incentive to minimize the time people spend on their facilities. After calculating these durations, defined as the difference between the last contact with a TfL-service and the first, with the exception of buses, we estimate the following regression:

$$\ln(duration_{it}) = \alpha_i + \beta \cdot d_t^{\text{post}} + \gamma \left(d_t^{\text{post}} \cdot d_i^{\text{treat}}\right) + \epsilon_{it.} \tag{2}$$

Note that our dependent variable is the natural logarithm of duration, so that coefficients can be interpreted as percentages. Once more, our main interest lies in the estimate of γ. Estimation results are shown in Table VI below, again for the five different characterizations of the treatment group. As can be seen from the table, our estimate of γ is consistently negative which suggests that commuters who were part of the treatment group cut their "time spent on public transport" by more than their non-treated peers. On average, the treatment group seems to cut their journey time by about 1% more. Given that the average journey in our sample lasts approximately 32 minutes, this amounts to an ex-ante average time-gain of about 20 seconds on a one-way commute for those who were treated by our preferred treatment measure. Since about 5% of treated commuters switch permanently, the measured average time gain for those commuters who actually switch is around 400 seconds (20 times 20).

Table VI: OLS-DiD results for commuter travel time.

	(1: different station)	(2: station strike)	(3a: factor 1.2)	(3b: factor 1.5)	(3c: factor 2)
	ln(duration$_{it}$)	ln(duration$_{it}$)	ln(duration$_{it}$)	ln(duration$_{it}$)	ln(duration$_{it}$)
β	0.00711***	0.00113	0.00125	0.000670	-0.000698
	(0.00164)	(0.00158)	(0.00132)	(0.00108)	(0.00103)
γ	-0.0124***	-0.00518**	-0.00548***	-0.00977***	-0.0121***
	(0.00206)	(0.00204)	(0.00198)	(0.00261)	(0.00430)
obs	312,103	312,103	312,103	312,103	312,103

Notes. Table VI reports OLS estimates of equation (2) where the dependent variable is log travel time for commuter i on day t. The columns present the results for each measure of treatment (defined in Box 1). β is the coefficient of the post-strike dummy (d_t^{post}) and γ is the coefficient of the interaction term between the post-strike and treatment group dummies ($d_t^{post} \cdot d_i^{treat}$). * denotes significance at the 10% level, ** implies significance at the 5% level, *** indicates significance at the 1% level. Individual fixed effects are applied in each estimation and robust standard errors are recorded in the parentheses.

7.4 Mechanism

Given that the previous sections have established that treated commuters were more likely to switch stations and cut travel time in the post-strike period than their non-treated peers, a logical follow-up question is: why? In the remainder of this section, we will provide evidence which suggests that this is due to the existence of informational imperfections. To make this point we use information on two characteristics of the London underground system that are not easily observed by commuters, namely map distortion (Section 7.4.1) and line speed (Section

7.4.2).

7.4.1 Map distortion

As noted before, an important source of imperfect information lies in the fact that the London Tube map provides a distorted picture of the geography of London. For the exercise in this subsection, we quantify these distortions in the following way: for each station on the map (S) we list those stations that lie within a 2km radius (which is about a 25 minute walk) from S. We subsequently correlate the true distance between these stations, with the distance on the Tube map, which we have digitized. Subtracting the resulting correlation from 1, gives our measure for distortion. Other measures where we take a 5km radius, or consider the closest 10 stations, give measures of distortion that correlate highly with the 2km-radius measure, with correlation coefficients of 0.96 and 0.94 respectively.

Map distortions are not constant across London: some people live in areas where the Tube map is more distorted than others, the general rule being that distortion increases with distance from central London. Thanks to this spatial variation, we are able to ask: do commuters who live in areas that are more distorted on the London Tube map, have greater difficulty in finding their preferred route? And do they learn more from the strike as a result? To answer this question, we estimated the following difference-in-difference-in-differences regression:

$$d_{it}^{\text{j\_mode}} = \alpha_i + \beta \cdot d_t^{\text{post}} + \gamma \left(d_t^{\text{post}} \cdot d_i^{\text{treat}} \right) + \zeta \left(d_t^{\text{post}} \cdot distortion_i^{\text{j}} \right) + \theta \left(d_t^{\text{post}} \cdot d_i^{\text{treat}} \cdot distortion_i^{\text{j}} \right) + \epsilon_{it}, \tag{3}$$

where "$distortion_i^{\text{j}}$" is our measure of map distortion around individual i's modal station of entry or exit (with j$\in$ {entry,exit}). Note that this exercise again explicitly distinguishes between the station of entry and exit, since map distortions are likely to be different at both ends. Table VII reports our results for γ and θ. In this regression, a negative estimate for θ would suggest that treated commuters who live in (or travel to) more distorted areas, are less likely to use their pre-strike modal journey in the post-strike period. This would provide evidence in favor of the hypothesis that commuters who live in more distorted areas, have greater difficulty in finding their optimal commute. And as can be seen from Table VII, this indeed seems to be the case: our estimate of θ tends to be significantly negative across specifications, thereby pointing towards the importance of informational imperfections in explaining our findings. The estimates of γ remain negative and significant in our preferred specification (row 1), which suggests that map distortion cannot explain the full effect, but the fact that the absolute value of our estimate tends to go down suggests that it does explain part of it.

Table VII: Estimates of γ and θ when interacting with map distortion.

Treatment definition	(1: γ)	(2: γ)	(3: θ)	(4: θ)
	Entry Mode	Exit Mode	Entry Mode	Exit Mode
1. different station	-0.0152***	-0.0407***	-0.0675**	-0.0263
	(0.00478)	(0.00661)	(0.0327)	(0.0438)
2. station strike	0.00927*	0.00278	-0.0996***	-0.0971**
	(0.00494)	(0.00677)	(0.0340)	(0.0455)
2a. entry on strike	0.0110**	-	-0.160***	-
	(0.00482)	-	(0.0376)	-
2b. exit on strike	-	0.00335	-	-0.142**
	-	(0.00838)	-	(0.0612)
3a. time factor(1.2)	-0.0103**	-0.0183***	-0.0111	0.0207
	(0.00490)	(0.00668)	(0.0333)	(0.0444)
3b. time factor(1.5)	0.000678	-0.0257***	-0.0712*	0.0540
	(0.00597)	(0.00793)	(0.0402)	(0.0525)
3c. time factor(2)	0.0224**	-0.0124	-0.208***	0.0109
	(0.00942)	(0.0188)	(0.0633)	(0.0784)

Notes. Table VII reports OLS estimates of γ and θ from equation (3) for each treatment definition (defined in Box 1) where the dependent variable indicates whether commuter i on day t traveled using their modal station. γ is the coefficient of the interaction term between the post-strike and treatment group dummies ($d_t^{post} \cdot d_i^{treat}$) and θ is the coefficient of the triple interaction term between the post-strike and treatment group dummies with the measure of map distortion ($d_t^{post} \cdot d_i^{treat} \cdot distortion_i^j$). Columns 1 and 3 report estimates where commuter i traveled on their modal entry station and columns 2 and 4 report estimates where commuter i traveled on their modal exit station. The number of observations is 267, 588, except for treatment definitions (2a) and (2b) where it is 226,404 and 184, 482 respectively.Table VII provides results from 12 separate estimations. The full results of each estimation can be found in Tables A1, A2, and A3 in the Appendix. * denotes significance at the 10% level, ** implies significance at the 5% level, *** indicates significance at the 1% level. Robust standard errors are recorded in the parentheses.

7.4.2 Line speed

Even if the London underground network were to adopt an undistorted Tube map, this still would not solve all informational problems. The reason is that many characteristics of various lines, such as crowdedness, nature of the follow-up journey to work, etc. remain unknown until that line is actually tried. One such characteristic that is easily quantified, is line speed. The average speed at which trains travel differs considerably across lines, from as low as 15 km/h for the Hammersmith & City-line to nearly 50 km/h for the Waterloo & City-line.[23] Consequently, two journeys that look equally far on an undistorted map, are still not equivalent if they are made in trains that travel at different speeds.

Table VIII therefore reports results that were obtained after estimating the following difference-in-difference-in-differences regression:

$$d_{it}^{\text{mode}} = \alpha_i + \beta \cdot d_t^{\text{post}} + \gamma \left(d_t^{\text{post}} \cdot d_i^{\text{treat}} \right) + \zeta \left(d_t^{\text{post}} \cdot speed_i \right) + \theta \left(d_t^{\text{post}} \cdot d_i^{\text{treat}} \cdot speed_i \right) + \epsilon_{it} \quad (4)$$

Since speed varies across lines, we now limit ourselves to the sample of commuters who stay on the same underground line for their entire commute (the same sample that was used in Column 2 of Table V). Consequently, our speed-variable becomes individual i-specific. The "same line"-restriction reduces sample size, as a result of which our estimates become less significant (like in Table V).

Our exercise suggests that treated individuals are more likely to change their journey in the post-strike period if they were commuting on a relatively slow line before the strike. Because this regression includes speed, which is inversely related to slowness, a *positive* estimate for θ now provides evidence in favor of the idea that switchers move away from slower lines. The reason seems to be that the episode of forced experimentation during the strike makes slow-line commuters aware of the fact that their usual train is rather slow-paced, which induces them to reconsider their options post-strike. This is again consistent with the hypothesis that informational imperfections drive our main results.

[23]Calculations are based upon TfL-information and contain the average speed attained by the various trains in between stations. Consequently, our measure is not distorted by the density of stations on a particular line, which is a characteristic that is easily observed from the Tube map.

Table VIII: Estimates of γ and θ when interacting with line speed.

Treatment definition	(1: γ)	(2: θ)
1. different station	-0.163***	0.210***
	(0.0392)	(0.0695)
2. station strike	-0.0645	0.0797
	(0.0502)	(0.0882)
3a. time factor(1.2)	-0.129***	0.207***
	(0.0423)	(0.0750)
3b. time factor(1.5)	-0.0874*	0.112
	(0.0463)	(0.0807)
3c. time factor(2)	-0.168**	0.233*
	(0.0773)	(0.134)
obs	47,052	47,052

Notes. Table VIII reports OLS estimates of γ and θ from equation (4) for each treatment definition (defined in Box 1), where the dependent variable indicates whether commuter$i$ on day$t$ traveled using their modal station (entry or exit). γ is the coefficient of the interaction term between the post-strike and treatment group dummies ($d_t^{post} \cdot d_i^{treat}$), and θ is the coefficient of the triple interaction term between the post-strike and treatment group dummies with the measure of line speed ($d_t^{post} \cdot d_i^{treat} \cdot speed_i$). Table VIII provides results from 5 separate estimations. The full results of each estimation can be found in Table A4 in the Appendix. * denotes significance at the 10% level, ** implies significance at the 5% level, *** indicates significance at the 1% level. Robust standard errors are recorded in the parentheses.

8 Interpretation

This paper presents evidence that a significant fraction of commuters in our dataset failed to optimize their journey due to the existence of informational imperfections. As a result, a disruption was able to bring about lasting changes in behavior, yielding associated time-gains. We consider two competing hypotheses that could explain our findings.

Under Hypothesis I, agents in our dataset were acting rationally and followed the optimal search rule, but due to the presence of search costs they aborted their exploration for the best alternative before they had found their global maximum. Along these lines, Aghion *et al.* (1991) formally show that following the optimal search strategy does not necessarily imply that the global maximum will be found. Using the language of Baumol and Quandt (1964), Hypothesis I implies that although agents were not maximizing, they were optimizing (i.e. behaving optimally given the existence of search costs).

Under Hypothesis II, on the other hand, agents were not adhering to the optimal search rule and experimented too little relative to the prescription of the standard-rational model. Using Baumol-Quandt terminology, this hypothesis implies that agents were neither maximizing nor optimizing.

To investigate which of these two hypotheses is in the best position to explain our results, we consider the optimal search strategy for this problem, following Gittins (1979) and Weitzman (1979). Using Weitzman's formulation and notation, the optimal strategy is to continue trying new alternatives until

$$c_i = e^{-rt_i} \int_z^{\infty} (x_i - z) dF_i(x_i) - (1 - e^{-rt_i})z, \tag{5}$$

where c_i is the cost of trying a new alternative i, r is the discount rate, and t_i is the time lag at which the value of a new alternative is learned. The parameter z is the present discounted value of the alternative that is currently chosen, while x_i represents the present discounted value of the most attractive unexplored alternative i. This value is distributed according to a c.d.f. $F_i(\cdot)$. As the value of an alternative route is learned soon, if not immediately after trying it, we advance with $t_i = 0$ (such that $e^{-rt_i} = 1$). Equation (5) then simplifies to

$$c_i = \int_z^{\infty} (x_i - z) dF_i(x_i). \tag{6}$$

Welfare gains or losses from finding a better route can come in many forms. Gains may include the discovery of a line that is less crowded, while losses include the physical comfort, crowding and the inconvenience and anxiety that may be generated by uncertainty and searching for the correct train, platform, or transfer (Cheng, 2010). Here we focus on the one element we can observe: time spent in TfL systems. Our preferred specification in Table VI suggests expected ex-ante time savings of 20 seconds per journey. This is a rough estimate, as it ignores the time spent walking or cycling to and from stations, which may correlate with observed travel time. We approximate the average daily welfare-gain $\int_{z^{daily}}^{\infty}(x_i^{daily} - z^{daily})dF_i(x_i)$ realized by commuters who found a quicker route thanks to the strike by setting it equal to the expected present discounted value of the expected average time-gain. We follow Small (2012) in using half the gross wage rate to value the cost of travel time, for London an hourly travel-time cost of \$11.60. In this calculation we furthermore work with an annual discount rate of 4% and assume that any gains last for a period of 3.3 years, given that average job tenure in the UK is 9 years and the average time that London households live in their home is 11.8 years.[24] (The expected value of the minimum of two independent draws, one from a uniform distribution with a length of 9 and one with a length of 11.8, is 3.3.) We factor in weekends and holidays. The implied search cost for the net present value of benefits is $c_i = \int_z^\infty (x_i - z) dF_i(x_i) \approx \$49$.[25] This number corresponds to more than 4 hours of commuting time by our assumptions on the cost of commuting.

There are at least five reasons that make this estimation conservative. First, this calculation ignores all the unmeasured advantages of the new alternative, which may be substantial. Second, observed effects are likely to be larger if agents had experimented voluntarily in a more predictable environment, as opposed to the less predictable strike environment. Third, our assumptions on parameters (such as discount rate) are conservative choices. Fourth, we use income-based estimates for the cost of commuting; welfare-based estimates of the cost of commuting tend to be much larger (Ahlfeldt et al., 2015). Fifth, we count the benefits to each commuter once every day, even though many of them will benefit from them more than once on many days, when they commute back on the improved route.

If commuters were adhering to the optimal search strategy prescribed by equation (6), this implies that the cost of trying the most attractive untried alternative would have to be greater than \$49. Or stated otherwise: under the assumption that our data were generated by optimizing searchers, one would have had to offer a commuter more than \$49 (equivalent to a one-off commuting time gain of over 4 hours) in order to induce him to try the most attractive untried alternative for just one day, after which he is free to go back to his status quo again. This strikes us as implausibly high and suggests that agents underestimate the value of experimentation and experiment too little as a result. Thus this result points towards Hypothesis II, which states that commuters experiment less than prescribed by the standard search model.

Concerning the efficiency of the underground network as a whole, we find that the net present value of time saved is larger than the time lost on the strike days. As set out in Section 7.1,

[24] Sources: http://www.ons.gov.uk/ons/dcp171778_385428.pdf reports the average gross yearly income in London as GBP 34,346. According to http://www.incometaxcalculator.org.uk, this gross income translates into a net income of GBP 26,176. We assume a work week of 37.5 hours and an exchange rate of 0.76 GBP for one USD. Job and house turnover rates from http://www.cipd.co.uk/binaries/megatrends_2013-job-turnover-slowed-down.pdf and https://www.gov.uk/government/statistics/english-housing-survey-2013-to-2014-household-report. Our calculations account for the fact that London has a significant rental sector, where turnover tends to be much higher.

[25] To illustrate the robustness of our finding, pushing the annual discount rate up from 4% to 20%, would only validate the rational search rule for $c_i \approx \$37$, which still seems implausibly high. Using the quasi-hyperbolic discount function $\left\{1, \varphi\delta, \varphi\delta^2, \varphi\delta^3, ...\right\}$ (where δ is the daily discount factor, which we base upon an annual discount rate of 4%) and setting φ equal to the Laibson, Repetto and Tobacman (2008)-estimate of 0.7, yields $c_i \approx \$32$, which corresponds to more than three hours of commuting.

only a subset of about 5% of commuters found a better route to work thanks to the strike. The remaining 95% did not make such a discovery. They only suffered from delays on February 5 and 6. Following the exercise above, we estimate that the average discounted value of time saved for the subset of treated commuters is around 4 hours per treated commuter. Around 70% of commuters were treated by this main treatment definition (commuters who deviated from their pre-strike modal station during the strike). This suggests that the average discounted value of the gain is over 3 hours per commuter. The average time loss for people who remained in the system on strike day was less than 2 minutes (as in Figure IV). Even commuters who deviated from their pre-strike modal station during the strike experienced delays of only 4.5 minutes per trip. This suggests that the net present value of time saved due to the strike far exceeds time lost on the strike days.

These findings suggest that agents in our dataset were experimenting less than prescribed by the standard-rational search model. This is consistent with laboratory evidence surveyed and reported in Anderson (2012), but to the best of our knowledge our study is the first to present evidence in favor of this hypothesis based upon detailed field data.

9 Conclusion

This paper provides evidence to suggest that a significant fraction of commuters in London fail to find their optimal route to work. In our preferred estimate, this concerns around 5% of commuters. This failure seems to be driven in part by noisy information, such as the geographic information provided by the distorted, stylized London Underground Map and the fact that different train lines travel at different speeds.

Because agents are seemingly stuck in suboptimal habits, an exogenously-imposed constraint such as the London Underground strike of February 2014 brought lasting changes in behavior. These results can be interpreted two broad ways. The first is that commuters are not searching in a way that is consistent with the standard-rational model. The second is that search costs are high, making it rational for commuters to stop searching for better alternatives whilst traveling along routes that are markedly suboptimal. Given the magnitude and prevalence of the time savings that we find, we consider that the observed behavior is unlikely to be rationalized by search costs.

Our calculations furthermore show that time-gains on the Tube subsequently achieved by those who switched route, outweigh the time-losses incurred by all commuters during the strike. Consequently, the strike was probably efficiency-enhancing along the lines of Kaldor-Hicks. In the context of the London Underground, this implies that commuters could be made better off if given encouragement to experiment. As partial closure of the network is a costly way to achieve this, different designs of the Underground Map and use of internet sites created to support journeys could aid commuters.

Given that many challenges faced by people and firms are more complex and less repetitive than the commuter-problem analyzed in this paper, our estimate of suboptimal habits may be a lower bound to the problem in other contexts.

Our findings illustrate that people might get stuck with suboptimal decisions because of under-experimentation. As a result, the imposition of constraints can improve long-run efficiency. These results highlight the importance of implementing occasional routine breaks.

10 References

Aghion, Philippe, Patrick Bolton, Christopher Harris, and Bruno Jullien, "Optimal Learning by Experimentation", *Review of Economic Studies*, 58 (1991), 621-654.

Aghion, Philippe, Mathias Dewatripont, and Patrick Rey, "Corporate Governance, Competition Policy and Industrial Policy", *European Economic Review*, 41 (1997), 797-805.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall", *Econometrica*, 83, (2015), 2127-2189.

Ambec, Stefan, Mark A. Cohen, Stewart Elgie, and Paul Lanoie, "The Porter Hypothesis at 20: Can Environmental Regulation Enhance Innovation and Competitiveness?", *Review of Environmental Economics and Policy*, 7 (2013), 2-22.

Anderson, Christopher M., "Ambiguity Aversion in Multi-Armed Bandit Problems", *Theory and Decision*, 72 (2012),15-33.

Anderson, Michael L., "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion", *American Economic Review*, 104 (2014), 2763-2796.

Baumol, William J., and Richard E. Quandt, "Rules of Thumb and Optimally Imperfect Decisions", *American Economic Review*, 54 (1964), 23-46.

Berman, Eli and Linda T.M. Bui, "Environmental Regulation and Productivity: Evidence from Oil Refineries", *Review of Economics and Statistics*, 83 (2001), 498-510.

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

Bolton, Patrick and Christopher Harris, “Strategic Experimentation”, *Econometrica*, 67 (1999), 349-374.

Caplin, Andrew, Mark Dean, and Daniel Martin, “Search and Satisficing”, *American Economic Review*, 101 (2011), 2899-2922.

Cheng, Yung Hsiang, “Exploring Passenger Anxiety Associated with Train Travel”, *Transportation* 37 (2010), 875-896.

Copeland, Brian R. and M. Scott Taylor, “Trade, Growth, and the Environment”, *Journal of Economic Literature*, 42 (2004), 7-71.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg, “Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model”, *American Economic Review*, 96 (2006),1043-1068.

Gittins, John C., “Bandit Processes and Dynamic Allocation Indices”, *Journal of the Royal Statistical Society*, 41 (1979), 48-177.

Goodwin, Phil B, “Habit and Hysteresis in Mode Choice”, *Urban Studies*, 14 (1977), 95-98.

Gray, Wayne B, ”The Cost of Regulation: OSHA, EPA, and the Productivity Slowdown”, *American Economic Review*, 77 (1987), 998-1006.

Guo, Zhan, “Mind the Map! The Impact of Transit Maps on Path Choice in Public Transit”, *Transportation Research Part A*, 45 (2011), 625-639.

Hicks, John R., *The Theory of Wages* (London: Macmillan, 1932)

Hodges, Andrew, *Alan Turing: The Enigma* (London: Random House, 2014)

Jaffe, Adam B. and Karen Palmer, “Environmental Regulation and Innovation: A Panel Data Study”, *Review of Economics and Statistics*, 79 (1997), 610-619.

Laibson, David, Andrea Repetto, and Jeremy Tobacman, “Estimating Discount Functions with Consumption Choices over the Lifecycle”, NBER Working Paper No. 13314, 2008.

Matejka, Filip and Alisdair McKay, “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model”, *American Economic Review*, 105 (2015), 272-298.

Morgan, Peter and Richard Manning, “Optimal Search”, *Econometrica*, 53 (1985), 923-944.

Palmer, Karen, Wallace E. Oates, and Paul R. Portney, “Tightening Environmental Standards: The Benefit-Cost or the No-Cost Paradigm?”, *Journal of Economic Perspectives*, 9 (1995), 119-132.

Porter, Michael E., “America’s Green Strategy”, *Scientific American*, 264 (1991), 168.

Porter, Michael E. and Claas van der Linde, “Toward a New Conception of the Environment-Competitiveness Relationship”, *Journal of Economic Perspectives*, 9 (1995), 97-118.

Roberts, Kevin and Martin L. Weitzman, “Funding Criteria for Research, Development, and Exploration Projects”, *Econometrica*, 49 (1981), 1261-1288.

Rothschild, Michael, “Searching for the Lowest Price When the Distribution of Prices Is Unknown”, *Journal of Political Economy*, 82 (1974a), 689-711.

– “A Two-armed Bandit Theory of Market Pricing”, *Journal of Economic Theory*, (1974b), 185-202.

Simon, Herbert A., “A Behavioral Model of Rational Choice”, *Quarterly Journal of Economics*, 69 (1955), 99-118.

Sims, Christopher A., “Implications of Rational Inattention”, *Journal of Monetary Economics*, 50 (2003), 665-690.

Schmalensee, Richard, “The Costs of Environmental Protection”, mimeo, MIT, 1993.

Small, Kenneth A., “Valuation of Travel Time”, *Economics of Transportation*, 1 (2012), 2-14.

Van Exel, N. Job. A, and Piet Rietveld, “Public Transport Strikes and Traveller Behaviour”, *Transport Policy*, 8 (2001), 237-246.

Weitzman, Martin L., “Optimal Search for the Best Alternative”, *Econometrica*, 47 (1979), 641-654.

Wood, Wendy and David T. Neal, “The Habitual Consumer”, *Journal of Consumer Psychology*, 19 (2009), 579-592.

Zhu, Shanjiang and David M. Levinson, “Disruptions to Transportation Networks: A Review”, in: David M. Levinson, Henry X. Liu, and Michael G.H. Bell (eds.), *Network Reliability in Practice* (New York, NY: Springer, 2011).

11 Appendix

Table A1: OLS-DiD results when interacting with map distortion and treatment group is identified as different station (1).

	(1)	(2)
	Entry Mode	Exit Mode
β: d_t^{post}	-0.00440	-0.00317
	(0.00364)	(0.00511)
ζ: $d_t^{\text{post}} \cdot dist_i^{\text{j}}$	0.00141	-0.0435
	(0.0250)	(0.0338)
γ: $d_t^{\text{post}} \cdot d_i^{\text{treat}}$	-0.0152***	-0.0407***
	(0.00478)	(0.00661)
θ: $d_t^{\text{post}} \cdot d_i^{\text{treat}} \cdot dist_i^{\text{j}}$	-0.0675**	-0.0263
	(0.0327)	(0.0438)
obs	267,588	267,588

Notes. Table A1 reports OLS estimates of equation (3) where the dependent variable indicates whether commuter i on day t traveled using their modal station. β is the coefficient of the post-strike dummy (d_t^{post}), ζ is the coefficient of the interaction term between the post-strike dummy and the measure of map distortion ($d_t^{post} \cdot distortion_i^j$), γ is the coefficient of the interaction term between the post-strike and treatment group dummy (different station) ($d_t^{post} \cdot d_i^{treat}$) and and θ is the coefficient of the triple interaction term between the post-strike and treatment group dummy with the measure of map distortion ($d_t^{post} \cdot d_i^{treat} \cdot distortion_i^j$). Column 1 reports estimates where commuter i traveled on their modal entry station and Column 2 reports estimates where commuter i traveled on their modal exit station. * denotes significance at the 10% level, ** implies significance at the 5% level, *** indicates significance at the 1% level. Individual fixed effects are applied in each estimation and robust standard errors are recorded in the parentheses.

Table A2: OLS-DiD results when interacting with map distortion and treatment group is identified as individuals station strike (2).

	(1: entry on strike)	(2: exit on strike)	(3: station strike)	(4: station strike)
	Entry Mode	Exit Mode	Entry Mode	Exit Mode
β: d_t^{post}	-0.0196***	-0.0186***	-0.0210***	-0.0350***
	(0.00272)	(0.00612)	(0.00359)	(0.00510)
ζ: $d_t^{\text{post}} \cdot dist_i^{\text{j}}$	0.00306	-0.115***	0.0103	-0.00363
	(0.0207)	(0.0431)	(0.0254)	(0.0350)
γ: $d_t^{\text{post}} \cdot d_i^{\text{treat}}$	0.0110**	0.00335	0.00927*	0.00278
	(0.00482)	(0.00838)	(0.00494)	(0.00677)
θ: $d_t^{\text{post}} \cdot d_i^{\text{treat}} \cdot dist_i^{\text{j}}$	-0.160***	-0.142**	-0.0996***	-0.0971**
	(0.0376)	(0.0612)	(0.0340)	(0.0455)
obs	226,404	184,482	267,588	267,588

Notes. Table A2 reports OLS estimates of equation (3) where the dependent variable indicates whether commuter i on day t traveled using their modal station. β is the coefficient of the post-strike dummy (d_t^{post}), ζ is the coefficient of the interaction term between the post-strike dummy and the measure of map distortion ($d_t^{post} \cdot distortion_i^j$), γ is the coefficient of the interaction term between the post-strike and treatment group dummy (station strike) ($d_t^{post} \cdot d_i^{treat}$) and and θ is the coefficient of the triple interaction term between the post-strike and treatment group dummy with the measure of map distortion ($d_t^{post} \cdot d_i^{treat} \cdot distortion_i^j$). Column 1 reports estimates where commuter i traveled on their modal entry station and the measure of treatment is if their modal entry station was closed during the strike. Column 2 reports estimates where commuter i traveled on their modal exit station and the measure of treatment is if their modal exit station was closed during the strike. Column 3 reports estimates where commuter i traveled on their modal entry station and the measure of treatment is if their modal entry or exit station was closed during the strike. Column 4 reports estimates where commuter i traveled on their modal exit station and the measure of treatment is if their modal entry or exit station was closed during the strike. * denotes significance at the 10% level, ** implies significance at the 5% level, *** indicates significance at the 1% level. Individual fixed effects are applied in each estimation and robust standard errors are recorded in the parentheses.

Table A3: OLS-DiD results when interacting with map distortion and treatment group is identified by time factor (3).

	(1: factor 1.2)	(2: factor 1.2)	(3: factor 1.5)	(4: factor 1.5)	(5: factor 2)	(6: factor 2)
	Entry Mode	Exit Mode	Entry Mode	Exit Mode	Entry Mode	Exit Mode
β: d_t^{post}	-0.0101***	-0.0230***	-0.0158***	-0.0265***	-0.0177***	-0.0317***
	(0.00342)	(0.00479)	(0.00278)	(0.00382)	(0.00256)	(0.00351)
ζ: $d_t^{\text{post}} \cdot dist_i^{\text{j}}$	-0.0428*	-0.0782**	-0.0323*	-0.0820***	-0.0316*	-0.0712***
	(0.0235)	(0.0320)	(0.0189)	(0.0254)	(0.0174)	(0.0233)
γ: $d_t^{\text{post}} \cdot d_i^{\text{treat}}$	-0.0103**	-0.0183***	0.000678	-0.0257***	0.0224**	-0.0124
	(0.00490)	(0.00668)	(0.00597)	(0.00793)	(0.00942)	(0.0188)
θ: $d_t^{\text{post}} \cdot d_i^{\text{treat}} \cdot dist_i^{\text{j}}$	-0.0111	0.0207	-0.0712*	0.0540	-0.208***	0.0109
	(0.0333)	(0.0444)	(0.0402)	(0.0525)	(0.0633)	(0.0784)
obs	267,588	267,588	267,588	267,588	267,588	267,588

Notes. Table A3 reports OLS estimates of equation (3) where the dependent variable indicates whether commuter i on day t traveled using their modal station. β is the coefficient of the post-strike dummy (d_t^{post}), ζ is the coefficient of the interaction term between the post-strike dummy and the measure of map distortion ($d_t^{post} \cdot distortion_i^j$), γ is the coefficient of the interaction term between the post-strike and treatment group dummy (time factors) ($d_t^{post} \cdot d_i^{treat}$) and and θ is the coefficient of the triple interaction term between the post-strike and treatment group dummy with the measure of map distortion ($d_t^{post} \cdot d_i^{treat} \cdot distortion_i^j$).Columns 1, 3, and 5 report estimates where commuter i traveled on their modal entry station. Columns 2, 4, and 6 report estimates where commuter i traveled on their modal exit station. * denotes significance at the 10% level, ** implies significance at the 5% level, * denotes significance at the 10% level, ** implies significance at the 5% level, *** indicates significance at the 1% level. Individual fixed effects are applied in each estimation and robust standard errors are recorded in the parentheses.

Table A4: OLS-DiD results for commuters using modal station when interacting with line speed.

	(1: not on mode)	(2: station strike)	(3a: factor 1.2)	(3b: factor 1.5)	(3c: factor 2)
	Mode Station	Mode Station	Mode Station	Mode Station	Mode Station
β: d_t^{post}	-0.0133	-0.0448	-0.0250	-0.0821***	-0.0894***
	(0.0286)	(0.0449)	(0.0335)	(0.0235)	(0.0210)
ζ: $d_t^{\text{post}} \cdot speed_i$	0.0196	0.0476	-0.00320	0.0977**	0.106***
	(0.0520)	(0.0786)	(0.0601)	(0.0418)	(0.0371)
γ: $d_t^{\text{post}} \cdot d_i^{\text{treat}}$	-0.163***	-0.0645	-0.129***	-0.0874*	-0.168**
	(0.0392)	(0.0502)	(0.0423)	(0.0463)	(0.0773)
θ: $d_t^{\text{post}} \cdot d_i^{\text{treat}} \cdot speed_i$	0.210***	0.0797	0.207***	0.112	0.233*
	(0.0695)	(0.0882)	(0.0750)	(0.0807)	(0.134)
obs	47,052	47,052	47,052	47,052	47,052

Notes. Table A4 reports OLS estimates of equation (4) where the dependent variable indicates whether commuter i on day t traveled using their modal station (entry or exit). β is the coefficient of the post-strike dummy (d_t^{post}), ζ is the coefficient of the interaction term between the post-strike dummy and the measure of line speed ($d_t^{post} \cdot speed_i^j$), γ is the coefficient of the interaction term between the post-strike and treatment group dummy ($d_t^{post} \cdot d_i^{treat}$) and and θ is the coefficient of the triple interaction term between the post-strike and treatment group dummy with the measure of line speed ($d_t^{post} \cdot d_i^{treat} \cdot speed_i^j$). The columns report estimates of each of the treatment measures (defined in Box 1). * denotes significance at the 10% level, ** implies significance at the 5% level, *** indicates significance at the 1% level. Individual fixed effects are applied in each estimation and robust standard errors are recorded in the parentheses.

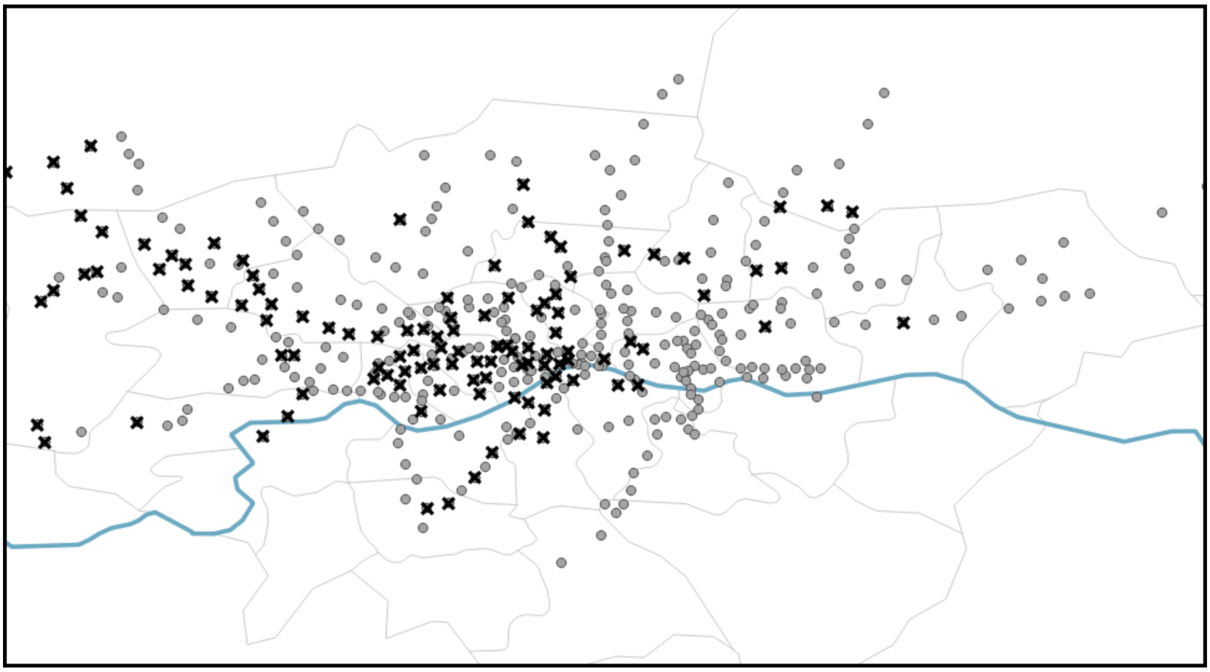

Figure 1

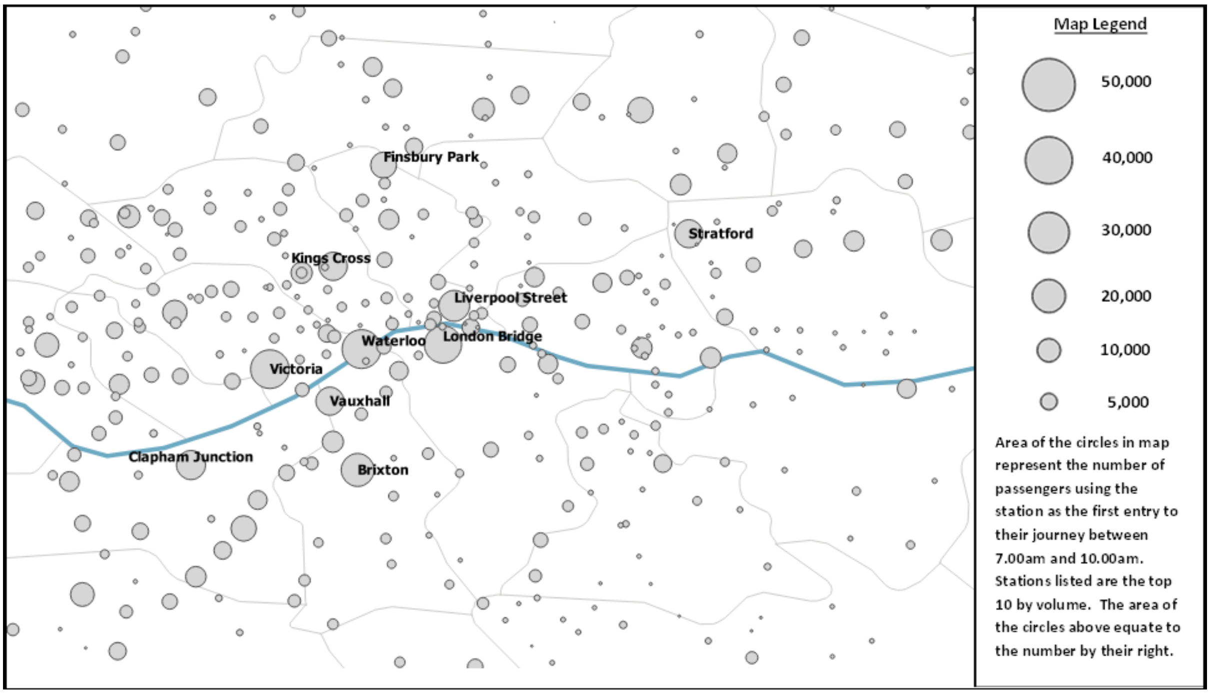

Figure 2a

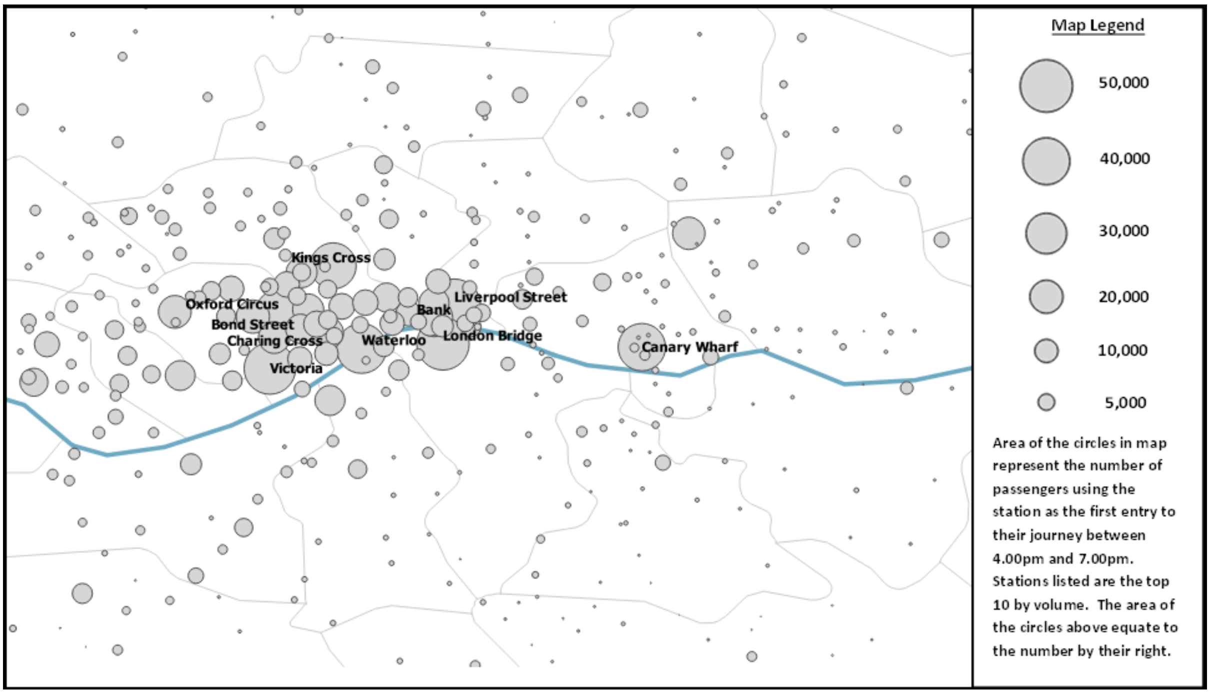

Figure 2b

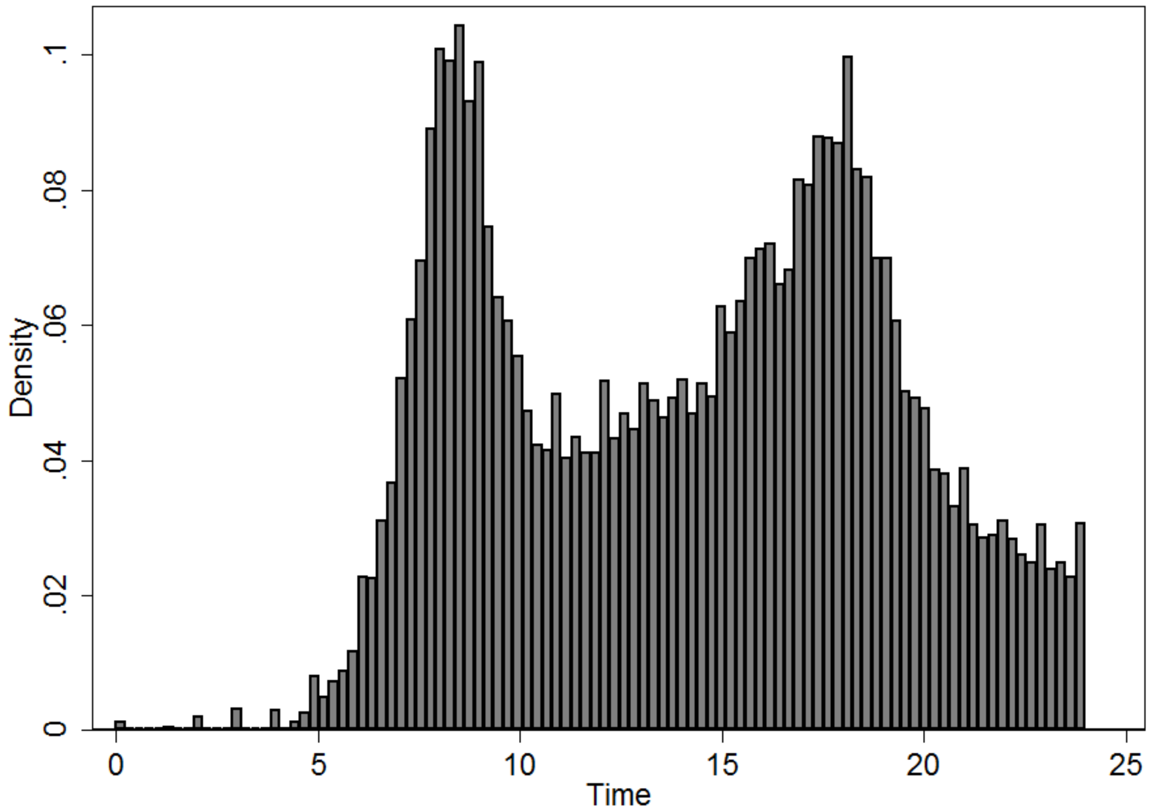

Figure 3

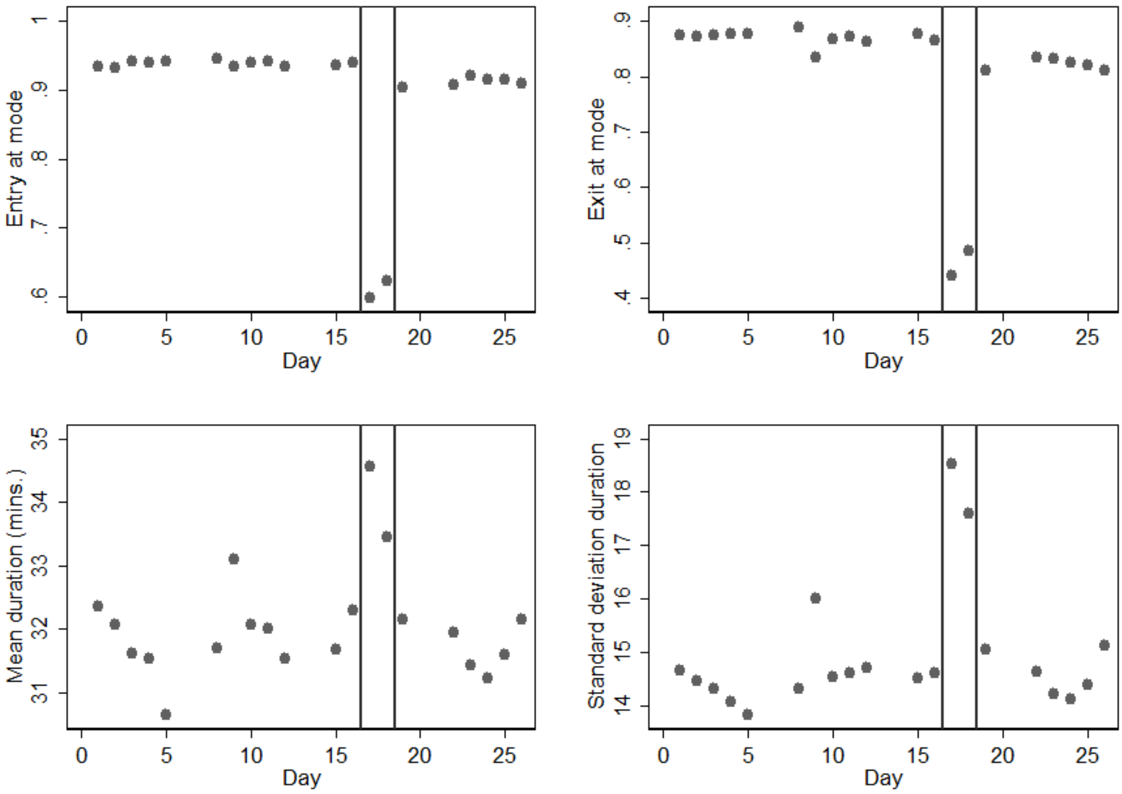
Entry at mode
.6
.7
.8
.9
1
0
5
10
15
20
25
Day
Exit at mode
.4
.5
.6
.7
.8
.9
Day
Mean duration (mins.)
31
32
33
34
35
Day
Standard deviation duration
14
15
16
17
18
19
Day

Figure 4

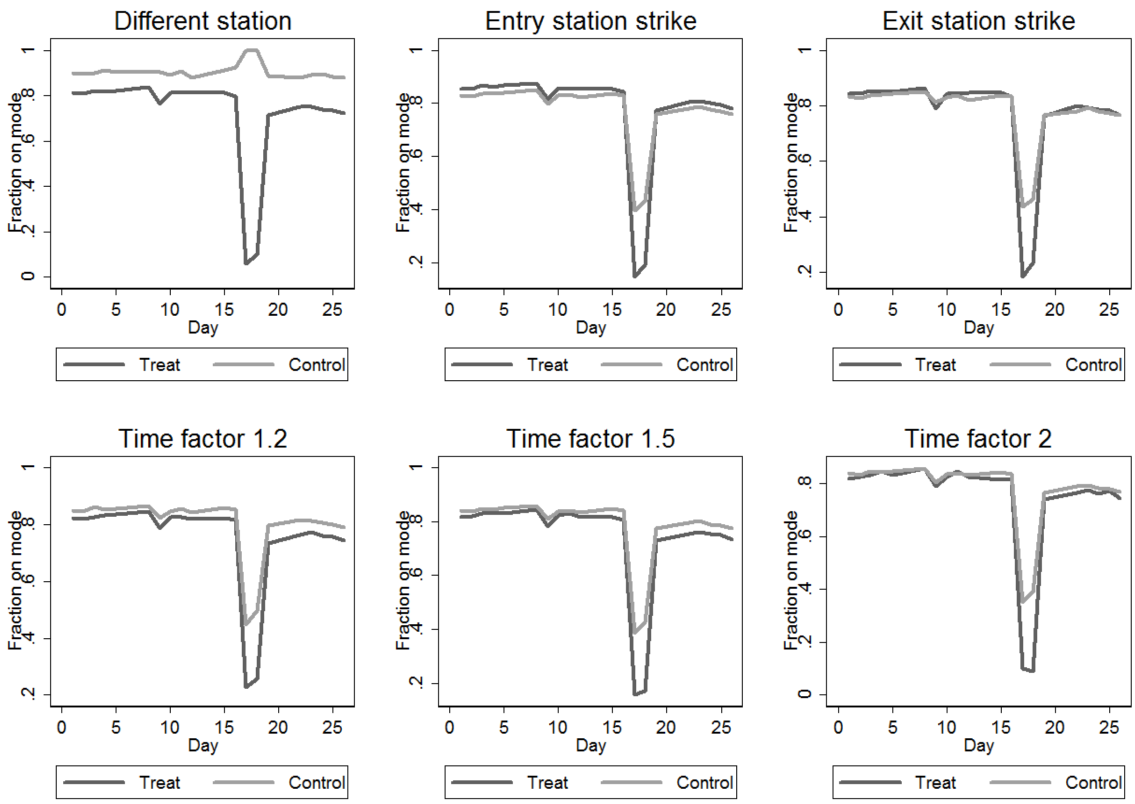
Different station
Entry station strike
Exit station strike
Time factor 1.2
Time factor 1.5
Time factor 2
Fraction on mode
Day
Treat
Control

Figure 5