How Smart is Your Smartcard? Measuring Travel Behaviours, Perceptions, and Incentives

Neal Lathia, Licia Capra
Department of Computer Science
University College London
Gower Street, London, UK
n.lathia, l.capra@cs.ucl.ac.uk

ABSTRACT

The widespread adoption of automated fare collection (AFC) systems by public transport authorities around the world means that, increasingly, people carry and use passive sensors (embedded inside of public transit tickets) to record their daily movements. Unlike mobile phones, the records held by AFC systems provide a rich and detailed source of data about peoples' transport habits: times of travel, modalities, destinations, trip durations, and fares paid. In this work, we explore the extent that this data offers the possibility to both build and measure future of travel-based ubiquitous computing applications. We focus on two potential end-users: first, how *travellers* may be aided by feedback mechanisms in order to re-align misperceptions of their travel behaviour and leverage this data to change their habits. In particular, we analyse differences between 85 travellers' surveyed perceptions of their public transport habits and their actual usage of the system. Second, how *transport authorities* can use this data to measure and implement incentive mechanisms that produce the expected impact. We use anonymised AFC data to measure the extent that financial incentives implemented by London's transport authority (such as peak-hour fares and student discounts) correlate with measurable changes in millions of travellers' behaviours.

Author Keywords

User Study, Incentive Mechanisms, Behavioural Awareness and Change

ACM Classification Keywords

H.4.m Information Systems Applications: Miscellaneous.

General Terms

Human Factors, Design, Measurement

INTRODUCTION

Many applications of pervasive computing use mobile phones or active sensors embedded into the environment in order to achieve their goals [1]; recent examples include using phones to encourage sustainable travel behaviours [2] and sensors to measure household water usage [3]. The advent and widespread adoption of automated fare collection (AFC) systems by public transport authorities around the world now means that people are also regular carriers of *passive sensors* that they may use when moving about urban spaces. These systems, based on radio-frequency identification (RFID) contactless "smart" cards (e.g., the Pasmo System in Tokyo, Japan, the Oyster card in London, England), not only store each traveller's current tickets, but generate streams of detailed travel records that allow transport operators to construct rich profiles of their passengers' habits [4, 5].

Travellers will often have complex and seemingly contradicting needs with regards to their daily travel habits, including temporal constraints (time to destination), flexibility, cost of travel, as well as being influenced by the availability of public transport and a varying willingness to engage in eco-friendly travel behaviour [2, 6]. By recording peoples' daily travel habits, smart card data implicitly captures how people balance their travel requirements, paving the way for systems that offer personalised [7] and live [8] travel information services. To date, however, this data has been studied in relation to the performance of the public transport system itself: for example, to estimate journey time or model transit demand [9, 10]. By being a system that links *individual* travellers to the public transport network, this data also has the potential to give insights into individuals' habits and test hypotheses relating to many aspects of urban mobility.

In this paper, we demonstrate how AFC records reveal hidden individuals' behaviours and responses to travel incentives. In particular, we use this data to evaluate two hypotheses, relating to two target audiences:

1. **Travellers' perceptions of their usage of public transport do not match their actual behaviour**. Are there any differences between people's perceived and actual transport-usage habits? This question is important for two reasons: (a) travellers will be basing their travel decisions on their perceptions; misperceptions may lead to incorrect decisions, and (b) feedback applications that record and visualise travellers' behaviours may leverage common misperceptions of travel to induce an awareness and encourage travellers to make positive changes in their habits.

UbiComp '11, Sep 17-Sep 21, 2011, Beijing, China.

In the first part of this paper, we report on the design of, and results from, a survey that, combined with data from smart cards, allows us to compare the differences between how 85 people report that they both purchase tickets and move in a city with how they actually do it. We believe that this is the first in-depth study to not only directly compare self-reported and actual public transport *usage*, but also the perceived relation between usage and payment.

2. **Transport operators offer incentives that do not work**. Public transport systems around the world will each have implicit incentives built into them; for example, they may implement time-varying fares (e.g., charging higher fares for rush-hour travel, to discourage non-essential travel). Do travellers respond to these incentives? We propose to test this question by using AFC data. An understanding of these aspects will be the foundation for any system that aims to stimulate additional behaviours.

   In the second part of this paper, we examine and enumerate a number of incentives that have been implemented by the public transport authority in London, England. We then seek to determine the extent that travellers are responding to these incentives by looking for evidence of the encouraged behaviour in two large, anonymised datasets of trips and ticket purchases of London travellers.

BACKGROUND: TRANSPORT FOR LONDON

The Transport for London (TfL) public transport infrastructure is a vast, multi-modal network of underground trains (11 interconnected lines with 270 stations), overground trains (5 lines with 78 stations) and buses (about 8,000 buses serving 19,000 stops) as well as trams, river services, and other specialised services. At the broadest level, travellers must opt to either use a single, contact-less smart card (the Oyster card) to pay for their journeys, or buy paper-based tickets. Travellers who use an Oyster card must decide between purchasing credit that is then deducted on a per-trip basis (i.e., using "pay as you go") or buying a travel pass (i.e., "travel card"). Pay as you go fares will vary based on the time of day, the locations of the origin and destination, and the travel modality. Travel cards, instead, are valid for unlimited travel within a certain geographic area, for a particular length of time (1 week, 1 month, 1 year), and sometimes for a specific transport modality (i.e., bus only). While these details are certainly specific to London, we note that other cities across the world share similarities. For example, New York's MetroCard allows for purchase of unlimited travel during 7-day and 30-day periods and San Francisco's BART's fares take into account length and speed of trip, as well as differentiating between user groups.

TRAVELLER SURVEY DESIGN

In order to understand how travellers *report* that they move around their city, we designed an online survey. The survey was composed of 25 questions covering two broad areas of interest: how respondents perceive their travel and how they perceive their fare purchasing habits. The first set of questions were about week day and week end travel habits: typical number of trips per day, travel time, modality of choice, whether trips are multi-modal, where their typical origin and destinations are, and how they rate the consistency of their habits (1 star was "very irregular" and 5 stars was "very regular"). The second set of questions dealt with purchasing habits instead. We split the questions into two groups: "when you use pay as you go," and "when you use travel cards." For the former, we asked *how much* credit they tend to purchase (if they ever do), and both *when* and *why* they chose to purchase credit. We also asked what travel card they bought (if they ever do), where they are valid, and why they opt for this purchase.

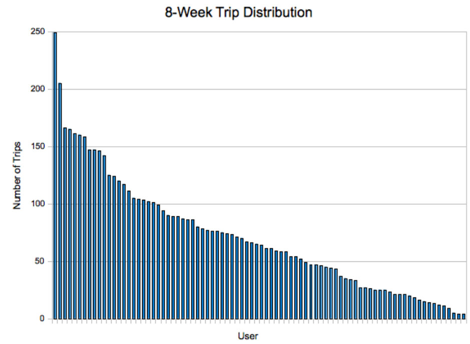
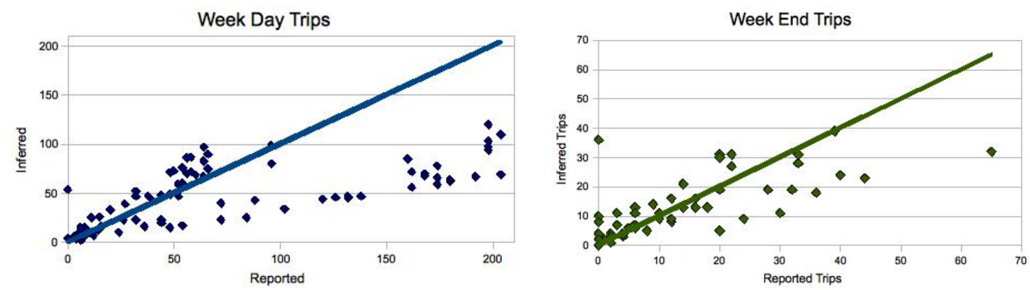


Figure 1. The user-trip distribution over 8-weeks of travel history; our respondents took between 4 and 249 trips.

SURVEY AND SMART CARD DATA COLLECTION

We disseminated the resulting web-form questionnaire online and received 119 responses. The survey data tells us about respondents' perceptions of their own mobility; the next step we took was to obtain the data of their actual trips. Transport for London currently stores eight weeks' travel history for each registered Oyster card; after the eight weeks, the data is anonymised. The last question of our survey thus asked for (a) permission to retrieve the respondent's travel history and (b) the unique identifier on their Oyster card that would enable us to do so. In total, 85 respondents allowed us to retrieve their history. Note that, once we received the data, we assigned a unique identifier to match travel history with survey responses and, in so doing, effectively re-anonymised the data (we cannot link it back to the individual users who responded to the survey). Furthermore, we explicitly did not collect any demographic information (other than the *type* of Oyster card the respondent owned) in order to comply with TfL's privacy requirements relating to disclosing the Oyster card histories. The Oyster card type gives us an indication of the demographic groups we captured in the survey: the majority of respondents (62%) owned adult Oyster cards, while 30% held 18+ student cards (required to have access to student fare discounts), 1% had 60+/Disabled "freedom" cards (which entitle the bearers to free travel), while the other 7% listed "other" or did not input an answer.

In Figure 1 we show the data's distribution of trips. In the data we received, which spans from 11 November 2010 to 1 January 2011, the respondents used their Oyster card between 4 and 249 times. This is an early indication of the widely varying requirements that city residents have with regards to their public transport network: some rely on it for



(a) Week Day Trips (b) Week End Trips

Figure 2. Comparing the survey results and smart card data on trips per day (left, on week days; right, on week ends)

continuous use, while others will be very infrequent users. In order to compensate for seasonal effects (i.e., Christmas and New Year), when travellers habits may be change significantly, we exclude any data after December 23rd from the following analysis.

Transport for London also provided us with two additional, fully anonymised, datasets. The first contains *all* Oyster card records from a one month period (March 2010). The second contains trip, payment and Oyster card details data from a 5% sample of travellers (roughly 300,000 people) over two 83-day periods: 3 May to 25 July 2009 and 18 October 2009 to 9 January 2010. These datasets allow us to compare the results from our survey respondents to the wider population. In the following sections, we refer to these as the 100% and 5% datasets respectively.

COMPARING PERCEIVED AND ACTUAL BEHAVIOUR

The first hypothesis that we test is whether travellers' perceptions of public transport usage match their actual usage of the system. In this section, we compare the *reported* behaviours from our survey to the smart-card recorded trips made by the survey respondents. We decompose the analysis into a number of sections: trips per day, travel times, choice of modality, geographic areas of interest, and fare purchasing habits. Following this comparison, we summarise the key findings and enumerate a number of opportunities for feedback-based systems that leverage this data.

Trips Per Day: Incorrectly Estimating Transport Usage

The first point on which we compare the survey and smart card data is the estimated trips per day. We defined a "trip" as a journey from one place to another using public transport. As an example, we described going from home to work and back home again as two trips. Respondents had to estimate how many trips they take in a typical week day and week end day (Saturday *or* Sunday), and were asked whether their typical trips are multi-modal. We then compute how many trips they have taken per day by looking at their Oyster history. To account for potential multi-modal trips, we only count a smart card swipe as a new trip if it is at least half an hour after the last trip's exit swipe. In Figure 2, we compare the reported and inferred values. In particular, we multiply the reported value (e.g., "2 trips per week day") by the number of (week day) travel days that we ob-

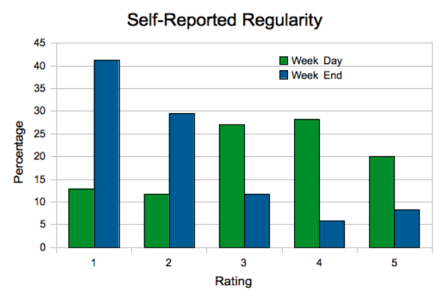
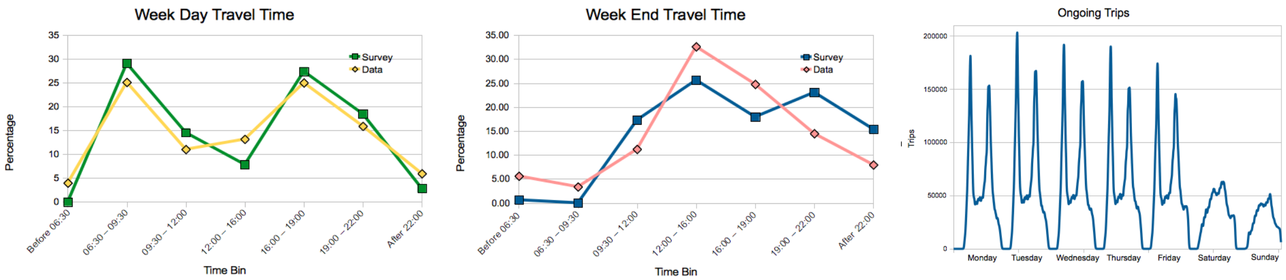


Figure 3. Comparing the survey respondents' self-evaluated trip regularity (where 1* is "very irregular" and 5* is "very regular") for week days and week ends.

serve in the data, and compare this result to the number of actual inferred trips from the data. In this case, making an accurate estimate of one's daily travel habits translates into a value that lies near the $y = x$ line (shown in each figure). We find that a large proportion of users consistently overestimate their usage of public transport: the *inferred* values are less than those that were *reported*. The distance between each point and the $y = x$ line shows the extent that travellers incorrectly estimate their travel habits. By normalising each distance by the number of days each user has travelled, we can get an intuition as to just how far estimation errors arise: in the week day data, travellers are off by 1.77 ± 1.51 trips; on week ends, this value lessens to 0.98 ± 0.88 per user.

Users were also given the option to claim that they never use public transport during week days: 2 of the 85 respondents opted for this answer. However, the smart card data held no evidence of this: all of the respondents used the public transport system in (at least) 1 week day. As per Figure 1, this may be explained by low-frequency travellers (for example, those who used public transport 4 times in the data). The differences between reported and computed average trips per day on week ends (Figure 2(b)) are not as high as the week day data. A total of 18 respondents claimed to not travel during week ends, and 17 smart card histories were missing week end trips. As before, many respondents tend to underestimate the number of trips they take, while overestimating that they only take 1 trip.

In this case, the data shows that, in general, respondents



(a) Week Day Trip Times (b) Week End Trip Times (c) 100% Dataset Avg Week

Figure 4. Comparing the survey results and smart card data on trip times per day (left, on week days; middle, on week ends) along with the number of ongoing trips in the anonymised dataset London travellers, clearly showing the commuting trends during week days.

are unable to accurately estimate how much they use public transport. Respondents were also asked to rate the extent that they believe their week day or week end trips are regular; where a 1 star rating means "very irregular" and a 5 star rating implies "very regular." The distributions of these ratings are shown in Figure 3. Overall, respondents tend to claim that their week day trips are regular (the distribution is biased toward the right), while week end trips are irregular (the distribution leans toward the left). This result contrasts the previous comparisons, where we found that travellers were better at estimating the frequency of their week end trips: it thus seems that travellers are better at estimating their irregular travel habits!

Travel Times and Peak-Time Commutes

The second topic covered in the survey asked about typical travel times. The available answers were split into ranges: before 06:30 AM, 06:30 to 09:30 AM, 09:30 to 12 PM, 12 to 4 PM, 4 to 7 PM, 7 to 10 PM, and after 10 PM. These bins include the two peak-fare times defined by TfL (06:30 to 09:30 AM and 4PM to 7PM), when week day travel, on a single fare, is more expensive. Unlike before, we found that, in this case, respondents' claims were closer to their actual data in the week days rather than week ends (Figures 4(a) and 4(b)). This is likely due to commitments relating to work: the two-spiked commuting pattern clearly emerges in both the survey and travel history data. However, travel outside of peak hours is underestimated; these may include, for example, opportunistic trips taken during lunch-time hours. A difference of a similar magnitude appears in the week end data (Figure 4(b)), where respondents report that they travel later than they actually do.

We also looked into the 100% Oyster card datasets to seek out any differences in the respondents of our survey and the wider population. In Figure 4(c), we show the cumulative ongoing trips over a week. The two-spiked commuting trend is very apparent in this data too, especially when compared to the pattern observed on week end days.

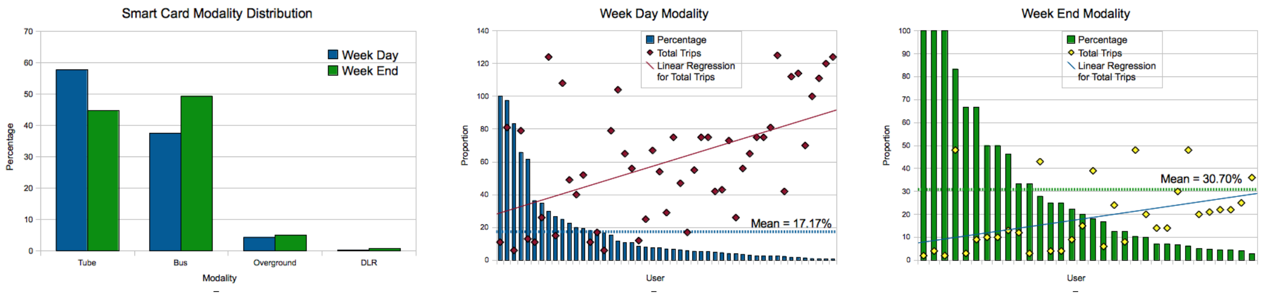
Travel Modality: Flexible or Opportunistic Trips

TfL operates a large multi-modal network. By turning to the smart card data and examining the distribution of travellers' trips with different transport modes (Figure 5(a)), we find that the underground (the "tube") is the most popular transport option, with over 50% of week day trips. However, there is a distinct shift on week ends, when buses carry a higher proportion of travellers than the underground instead (although, compared to week days, there are fewer trips).

The survey included a question of typical transport modality. In this case, we can use the travel modes that people claim to use on typical days to infer trips that may fall outside of their routine, and quantify how much of their travel history represents a-typical trips. We first examined the proportion of users who did not claim to use a particular modality, but actually did during their Oyster history. For example, we looked for respondents who reported to only use the underground during the week, but had week day bus trips registered in their travel history. Of the 85 total users, 47 (55.3%) used modalities that they did not claim to typically use during week days. On week ends, this number fell to 27 (31.8%).

An intuitive explanation of these numbers would be that respondents are telling us what they "typically" do, and this mismatch may arise from irregular, off-the-beaten track trips. If this were the case, we would expect that the proportion of these trips would be low, compared to all the trips by each traveller. In other words, if a person claims to not ride the bus during the week, but has done so, we would expect that the proportion of week day bus trips for this user would be small compared to the total number of trips taken. We therefore computed the distribution of per-user trip proportions that are taken on modalities that they claim to *not* use. The resulting distributions are in Figures 5(b) and 5(c); the corresponding means are 17.17% for week days and 30.70% for week ends. This means that, in our data, respondents take an average of 30.70% of all their week end trips on modalities that they did not claim to normally use on weekends. Both of the figures show a power-law distribution; most notably, there are some users who have taken more than half of their trips on modalities that they did not report to use.

There are two points to note here. First, the travellers with a very high proportion of these trips have a varying number of trips in their history (i.e., some of them are irregular users of public transport). Figures 5(b) and 5(c) include points that



(a) Smart Card Modality Distribution (b) Week Day (c) Week End

Figure 5. Usage of transport modalities: (a) the week day and week end distributions of trips on the underground, buses, overground rail and the Docklands Light Railway (DLR) system based on smart card data, (b) the distribution of the proportion of respondent's week day trips that they spend on modalities they do not typically use, along with the number of trips their smart card data contains, and (c) the same for week end data.

denote the total number of trips that each user has. The plots include a trend line, which is a linear regression on the total trips data. Both trend lines have a positive slope: as users have a higher number of trips, the proportion of trips that occur on modalities that they do not report to use decreases. However, there are exceptions to this trend. The dataset includes travellers who have both a high number of historical trips and a high proportion of trips on unreported modalities.

Origins, Destinations, and a-Typical Stations

The survey also asked respondents to list their rail trips' typical origins and destinations (the form itself allowed for an unlimited number of stations to be listed). By comparing these station names to the recorded smart card data, we can discover the extent that users travel between a fixed number of locations. As above, we are interested in comparing two variables: the proportion of trips that include an origin or destination that is *not* listed as a typical place and the total number of trips taken by that user. We note that two respondents used generic terms (e.g., "various," "others") when listing their week end destinations. One respondent even appended a comment to a station name: "when open," reflecting that this station is often closed during week ends due to ongoing construction work. Since we had no means of mapping generic terms to specific locations, we disregarded these users from the analysis.

On average, 64.09% of each respondent's week day rail trips were from or to a station that they did not list as a typical location. On week ends, that average rises to 87.86%. As above, the total number of trips in each user's smart card data is inversely proportional to percentage of user trips that include a-typical origins or destinations. In other words, the survey respondents who were more frequent users of public transport were better at enumerating the stations that they use. The adoption of generic terms further indicates that the respondents themselves seem to be aware of the fact that week end mobility is less likely to be between familiar locations, as they are not constrained by work-related commitments, and more likely to be disrupted by station closures. However, the average week day trips per user that include a-typical stations remains high. There are two types of trips that may explain this trend: (a) ad-hoc, opportunistic trips that travellers do not consider to be typical, and (b) adaptive travel behaviours (e.g., alighting at an earlier station and walking) when responding to congestion, station closures, or unforeseen events.

Cash-Fare Purchasing Habits: Overestimating Cost

There is currently no system for travellers to relate the way they use public transport to their fare purchasing habits [11]. Recall that travellers can opt to pay for their trips on a per-trip basis by buying travel credit (using *"pay as you go"* fares), or use 7-day, monthly, or annual passes, also called *travel cards*. The act of adding pay as you go credit to an Oyster card is denoted as "topping up," and travellers may do so online, via ticket machines or attendants in stations, or in local shops. In this section, we focus on these cash-fares that are bought by travellers.

In the survey, we asked users how much they typically top-up by each time they do so. The largest groups claimed that they spent between GBP 5 and 10 (33.78%), and between GBP10 and 20 (35.14%) at each purchase. The smart card data contained all the top-up amounts of our respondents; using this, we can compare the amounts that travellers *think* they typically spend at each purchase and how much they actually do, on average, spend each time. The data contained 381 purchases; in Figure 6 we compare the range that users claim to purchase within at each purchase and the actual average of all their purchases. Overall, majority of the

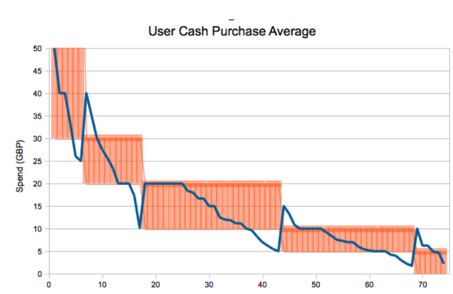


Figure 6. (Blocks) The ranges that users claim to spend when they top up and (line) their smart card average cash purchase: 20.27% of these users claim to spend more than they do.

averages fall within the range claimed by the respondent; we find that 10.81% spend, on average, more than they claim to. However, more importantly, 20.27% of respondents actually spend, on average, less than they claim: the average amount they spend per purchase is less than they claim. This is a substantial group of users who spend less on public transport than they think.

The 5% dataset also contained all of the top-up amounts of the sample of travellers, amounting to 3,379,570 top-up transactions. In these, we found that top-ups of less than GBP 5 accounted for 49.8% of all transactions, and 24.2% were between GBP 5 and 10. The minimum credit card transaction at ticketing machines is GBP 5 (which is also the smallest bank note value): these are thus top-ups made using coins. These figures show a much higher proportion of very small transactions than both our survey and respondents' Oyster histories showed, which reflects on the socio-economic status of the respondents we were able to reach (who were mostly university staff and students–people who have access to the web).

Summary: Opportunities for Feedback Applications

The analysis above displayed a wide variety of traveller behaviours. Differences emerge between travellers (reflecting differences between individuals' travel requirements) and between what survey respondents reported and actually did (misperceptions of travel patterns). In this section, we summarise our findings and relate them to potential applications that may be built from passively collected transport data.

Public Transport Usage Awareness. We noted that, while most respondents are using public transport less than they claim (Figures 2(a) and 2(b)), some clearly over-estimate how much they do, by reporting they take over five trips on a typical day. These results complement research that relate giving feedback to the uptake of public transport: behavioural awareness, in some cases, lead to an increase in public transport by 50% [12].

Regularity of Travel and Real-Time Information. Majority of respondents rated their weekend trips as highly irregular, but provided better estimates of how many trips they take on these days. Regularity thus seems to be a notion that is associated with time of travel (Figures 4(a) and 4(b)) and destinations rather than the amount of travel itself. This result could be used to augment research on real-time travel alerts, where feedback has been shown to reduce waiting time and increase travellers' satisfaction with public transit [13].

Journey Planning by Context. The increased usage of buses during weekends (Figure 5(a)) may indicate that buses are viewed as a less reliable means of transport, thus suitable for leisure but not for commuting (though other explanations could be applicable here). Also, 17.17% of respondents' week day trips and 30.70% of their week end trips were taken on modalities that they did not report to typically use. This result may indicate that travellers are more flexible in their travel habits than they report to be. The clear distinction in modality choice over different time periods is a key piece of information that AFC data can feed into trip planning systems.

Exploring the City. More than half of respondents' week day rail trips included a station that was not listed amongst their typical origins or destinations. Mobility data like this can be used to build recommender systems for city residents, letting them discover locations and social events that may be of interest to them, both in areas that they commonly frequent and unexplored parts of the city [14].

Fare Purchase Awareness. When considering only credit-based purchases, 20.27% of respondents claim to spend more than they actually do. Note that this statistic says nothing about those who buy the incorrect fare (where there was a cheaper option) and do not realise how much they might have saved. Building systems that facilitate the cost-effective use of public transport may also attract those travellers whose primary concern is how much they need to pay [11].

CORRELATING INCENTIVES AND BEHAVIOUR

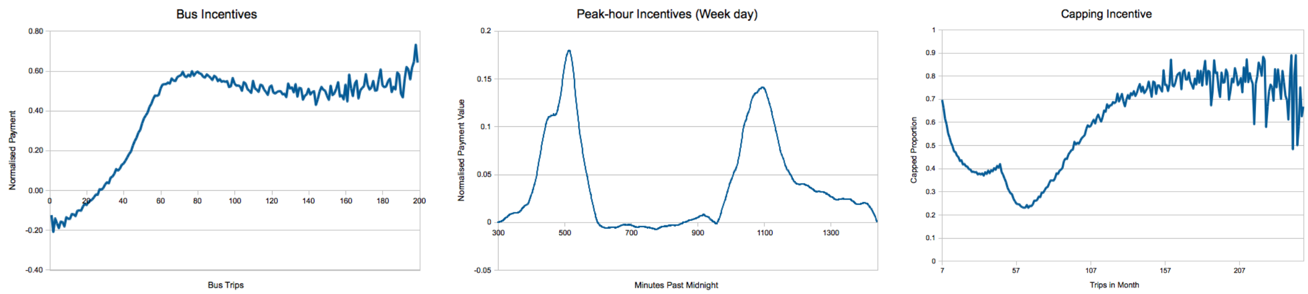
In the above, we focused on how travellers perceive their behaviour. We now turn to an alternative: how transport operators try to influence travellers' decisions. Travellers' behavioural choices are often guided by conflicting requirements: for example, the need to be at work on time vs. the choice of the cheapest modality; these choices may be swayed in a particular direction by providing different forms of incentives. Much like airlines, transport operators could give "reward miles" to loyal customers that can later be exchanged for free travel; they could also build games and design dynamic fare systems that encourage travellers to make use of their services. Incentives in the public transport domain are already well established: for example, bike sharing users in Paris, France, are given credit for taking bicycles to up-hill stations[1]. In this section, we show how AFC data can also highlight the incentives that travellers are responding to, and those that have no effect. In particular, we focus on the relation between usage and a number of features of the pricing system in London that are aimed to trigger particular travel and purchase behaviours.

Travel Cards: Encouraging Public Transport Usage

Oyster card users can choose to pay for public transport in a variety of ways. As above, the first decision is whether to purchase credit that will then be docked from their smart card on a per-trip basis (pay as you go) or buy a fixed-price pass that allows unlimited travel within a certain geographic area over a specified amount of time (a travel card). Travel cards may also be solely for use on buses or not restricted by modality at all.

The most generic difference between pay as you go and travel cards is that the former entails multiple (incremental) payments, while the latter requires a single payment followed by unlimited use. Since an important concern for travellers who use public transport is cost [2], the implications seems to be that the purchase of a travel card not only reduces the

[1] http://blog.velib.paris.fr/blog/?p=318



(a) Bus Usage and Payment (b) Week Day Payment Groups (c) Effect of Price Capping

Figure 7. Examining the effect of pricing incentives: positive values indicate a higher proportion of travel card passengers, while negative values represent a higher proportion of pay as you go passengers. Figure (a) shows that travel card owners will use buses more, Figure (b) reproduces the 2-spike commuting pattern and shows that pay as you go passengers tend to avoid peak-times, Figure (c) shows that pay as you go travellers who are given free travel (capped) will travel more.

barrier to entry into the system (since, throughout the validity of the travel card, no further financial commitments need to be made) but will also act as an implicit incentive to use the system, in order for travellers to feel that their initial payment is well spent. Does a fixed fare, with unlimited travel, encourage travellers to use public transport more?

To examine whether this incentive is reflected in the Oyster data, we focus on bus usage. Buses are not the most popular form of transit, and since they can be used for opportunistic trips (where even walking may be a viable alternative), they are most likely to show whether travel cards correlate with a higher use of the system. Unfortunately, since buses only require travellers to use their smart card upon entry, we cannot know how far each person travelled. The 100% dataset contains a total of 172,286,012 smart card bus trips. We pruned all of those which were taken with bus-only passes, leaving 115,173,544 trips. If travel cards were *not* an incentive to use buses, then we would expect travel card holders to use buses as much as non-travel card holders. More formally, we were able to quantify, for each traveller, how many trips she/he took in that month, and how many of them were taken using pay as you go or a travel card. We binned the data by number of trips; for each number of trips taken in the month i, we have $N_{i,PAYG}$ travellers on pay as you go and $N_{i,TC}$ travellers on travel card. We combined each pair of values as follows:

$$V_i = \frac{N_{i,TC} - N_{i,PAYG}}{N_{i,TC} + N_{i,PAYG}} \quad (1)$$

The normalised values V_i will be -1 when all of the travellers are on pay as you go, $+1$ if all the travellers are using travel cards, and 0 if there is an equal number of each. If travel cards were not an incentive to use buses, then we would expect all of these normalised values to be zero. The results, comparing number of bus trips taken in the month to the normalised value, is shown in Figure 7(a). Not only are none of the values zero, but there is a clear inflection point (from pay as you go to travel cards) at 28 trips in the month (an average of just below 1 bus trip per day): after this point, the use of travel cards vastly exceeds the use of pay as you go on the bus network.

Overall, in the data, 61% of bus trips and 75% of rail trips are completed using travel cards. This is both evidence that per-trip payment reduces the usage of public transport and that low-frequency users of public transport will opt to pay per-trip.

Peak Times: Discouraging Non-Commuters

The public transport authority in London has also implemented time-varying fares: during week days, between 06:30 to 09:30 and 16:00 to 19:00, pay as you go fares are more expensive (on rail services only). The change in price only affects pay as you go users; those commuters who are travelling with travel cards are exempt from fare differences between the peak and off-peak hours. The explicit rule is aimed to act as in incentive for non-essential travel to be avoided, in order to reduce congestion; this may couple with payment since those who are less frequent travellers will (or potentially should) be using pay as you go. However, as Figure 4(c) shows, the spikes in congestion remain unaffected by this rule. Is there any evidence to show that non-essential travel is being avoided (i.e., that this incentive works)?

As above, we can investigate the extent the smart card data holds evidence of the effect of this incentive by grouping travellers. The dataset contains 78,005,747 week day rail trips. From these, we compute two vectors, counting the normalised ongoing trips with travel cards and with pay as you go:

$$T_i = \frac{N_{i,TC}}{max(N_{TC})}; P_i = \frac{N_{i,PAYG}}{max(N_{PAYG})} \quad (2)$$

In this case, we performed the normalisation step first to account for the higher proportion of travel card trips that are taking place. We then combine each pair, in order to see the relative bias towards one fare type (pay as you go or travel card) over the course of a day:

$$V_i = T_i - P_i \quad (3)$$

As above, V_i will be positive when all trips are taken using travel cards and negative when there is a majority of pay as you go being used. If the variable fare structure did not sway travellers' decisions, then we would expect the combined and normalised distribution to be close to zero at all times (i.e., the balance between pay as you go and travel cards would be roughly the same at all times).

First, recall that the two-spiked commuting pattern, the number of ongoing trips over time, was displayed in Figure 4(c). The unnormalised vectors did already show that there are more trips outside of peak hours using pay as you go. The opposite is true for peak hours. After combining and normalising the data, shown in Figures 7(b) (week days), this pattern is highlighted even more clearly: by looking at the relative distribution over time of the fares that travellers use to access public transport, which have, built into them, incentives for peak and off-peak travel, we reproduce two-spiked commuting pattern. The travel data clearly shows that peak-fare pricing does not prevent peak-hours from being the most congested: what these results show is that peak-fare pricing, instead, guides travellers' purchasing decisions rather than their choice to travel.

Daily Capping: Switching to Free Travel

Pay as you go Oyster card usage is also subject to "daily capping:" the amount of credit that will be deducted from the traveller's smart card will not exceed the equivalent price of buying a daily travel card. In other words, once the capping threshold has been reached, travel throughout the rest of the day is free. This switch from paid to free travel should inherently act as an incentive to use public transport. For example, travellers may opt to use a bus (instead of a car, or even instead of walking) for short, opportunistic trips once they have already reached their daily cap.

The 100% dataset contains 91,391,155 trips taken using pay as you go. Of these, 90.37% are uncapped trips; travellers in this fare group often do not reach the capping limit. However, if their behaviour were unchanged when they do, we would expect to see a similar distribution of modality usage across both groups. In the uncapped group, 54.83% of trips are taken with buses. However, if we focus on the capped group, the proportion of bus trips rises to 73.53%. Suddenly, the usage of buses has become much more attractive.

We can examine this further by comparing the number of pay as you go trips that a traveller makes throughout the month with the proportion of those trips that have been capped (and are thus free). We consider *proportions* of trips since travellers need to pay until they reach the cap. Therefore, if they never do reach this cap (even if they still travel often) then a low proportion of their trips will be free. By averaging across all users we can see the trend that emerges across the city. The results are shown in Figure 7(c), where the x-axis denotes number of trips taken in the month and the y-axis is the average proportion of capped trips. The pay as you go travellers who are using the system the most are those who are taking advantage of the free travel they have earned.

Students Do Not Purchase Discounted Fares

The 5% sample data we have contains 184,923 students, identified by the type of Oyster card they use (note that TfL has no requirement that students must have student cards).

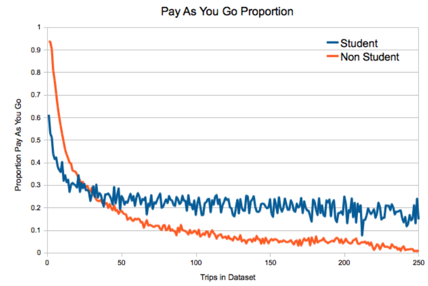


Figure 8. Comparing the number of trips taken in the dataset to the proportions of trips by students and non-students that were made using pay as you go. Although students are offered discounts for travel cards, even those who travel frequently use travel cards less than other travellers.

This card is valid throughout their period of study, and allows them to purchase travel cards at a discounted price (pay as you go is not discounted). We have a total of 10,723,473 trips taken by the students, and 48,888,268 trips taken by everyone else. If students were unaffected by the prospect of a discount for their travel card purchases, we would expect the relative proportion of trips taken with each fare type to be similar. However, we find that students slightly favour pay as you go trips: 35.15% of their trips were credit-based, compared to 31.88% of the pay as you go trips taken by everyone else (a difference of 3.27%). The data seems to show that students are not attracted to buying discounted travel cards.

We can deepen this analysis by comparing students and non-students who have the same travel requirements, or make the same number of trips in the dataset's time period. As before, if two people have the same high levels of travel requirements (expressed as number of trips in a given period) and one is a student, the availability of the discount would seem to imply that the student will be more likely to be using a travel card. In Figure 8, we find that the opposite is true. By plotting number of trips to average per-traveller proportion of trips using pay as you go, we find that the students who take a large number of trips consistently average more pay as you go trips than everyone else.

The data thus suggests that students are, overall, not benefitting from the discounts they are offered. There are a number of potential explanation for this. For example, purchasing a travel card requires a larger one-off payment than pay as you go (even though, cumulatively, more may be spent on pay as you go), and students may not feel to be in a financial position to commit to such a payment. The purchase of a travel card may be unappealing for other reasons too: one survey respondent commented that "as a student with a flexible timetable, it isn't always obvious whether a travel card is a better deal than pay as you go," while another stated that "my travel habits change every semester depending on courses."

Summary: Evidence of Financial Incentives

In this section, we have observed how smart cards can measure the effect of incentives given to travellers. All of the incentives operated by TfL are financial: they balance a ticket type (which has a cost) with an intended behaviour. However, we did not find that the former (ticket type or cost) always affects the latter (travel pattern); in fact, in some cases, the opposite is true. Furthermore, a number of relations emerged:

Accessibility and Balancing Requirements. By allowing passengers to easily access the system (with a one-off payment followed by unlimited travel), travel cards encourage higher use of the system: this is particularly noticed by the significantly higher proportion of bus trips that are taken by these travellers (note that bus stops do not have facilities—like train stations—to recharge Oyster credit).

Free Travel. As expected, those travellers who use pay as you go but reach the daily capping limit (after which travel is no longer charged for) are encouraged to, overall, use the system more. This relates to research by Shampanier, Mazar, and Ariely [15], who looked at the (often drastic) effect that free, or the appearance of free, services have on the purchasing options of customers.

Targeted Discounts. We found that not all students benefit from discounts — due to the fact that the discounted fares (travel cards) either do not match their travel requirements, which are often infrequent, or their budget of available funds. This observation once again highlights the relative importance of various travel requirements to different travellers (and demographic groups).

DISCUSSION

The cornerstone of the analysis above is the rising adoption of automated fare collection systems by public transport authorities, which translates into an ever-increasing number of travellers who carry passive sensors that record their travel behaviours. Although our focus was on London (Oyster), similar systems are in place in Beijing (Yikatong), Hong Kong (Octopus), Washington D.C. (SmarTrip), and Seattle (Orca), as well as many other cities around the world. Each system is designed in a unique way, both in terms of fare structure and geographic layout. However, the homogenising characteristic–the use of contact-less smart-cards–means that all of these systems will be subject to similar analysis.

Although the main task of these smart cards has been to act as a sensing device, to facilitate public transport access and payment, these powerful sources of data now also have the potential to become integral components of technologies aimed at *two* different audiences: towards both the travellers and transport operators. By being a source of long-term, individual public transport usage data for an entire city of commuters, these smart cards are powerful means of measuring how people use public transport (to aide future systems' design), how their perceptions reflect their actual usage (to build feedback applications), and the extent that they respond to incentives that are designed to encourage public transport usage (to measure the effectiveness of policy). In the future, as means for secure access to this data are enabled, a variety of applications may be designed to leverage it and help and promote urban mobility with public transport. The effect of any introduced incentive will be quantifiable using similar analysis on smart card data.

There are a number of dimensions of urban mobility that we did not cover in this work and leave as future work; for example, we did not investigate how users respond to or seek travel information when disruptions or delays occur. London also implemented a road congestion charge in 2003 (and extended it in 2007): drivers must pay a fixed fare to drive into the centre of the city between 7AM and 6PM during week days. Our Oyster data is from 2009 and 2010: earlier samples of the data would indeed show whether drivers were opting to leave their cars at home in exchange for a commute by public transport. However, an important caveat remains: reducing road congestion will eventually encourage travellers to return to their own vehicles [16]. The introduction of shared bicycles in cities around the world[2] now presents the opportunity to also build applications that intertwine exercise, health, and mobility using public transport.

RELATED WORK

Mobility has been a central theme of research into ubiquitous systems; an important theme in this area has been detecting mobility and inferring human activity from mobile phone data [17, 18, 19]. This data can then be leveraged to, for example, build social event recommender systems [14] or understand the design of a city via travel flows [20]. In this paper, we argue that AFC data is equally important for the design of future systems that will guide travel behaviour, manage travellers' individual needs, and provide incentives for sustainable transport. AFC data has the added benefit of not requiring an inference algorithm to operate on it (in order to, for example, classify transport modality [19]): the data itself contains fine grained details of travellers' behaviours.

Recent work has also examined how blind and deaf-blind people use public transit systems [21]. AFC records are also a rich, and non-invasive source of data about how this particular target group (along with all city residents who are ineligible for car licenses) use public transport[3].

The disparity between survey respondents' reports and their actual actions extends beyond transport contexts. Recently, similar investigations into shoppers' behaviours highlighted that shoppers claim to be more ethical in their purchasing choices than they actually are [22]. Our work is also part of a growing literature on voluntary behavioural change relating to public transport usage [23], including mobile-based solutions that visually represent the sustainability of travellers' choices [2, 24]. Mankoff et al. [25] have also investigated the role that social networks play in moulding peoples' decisions. The most direct form of incentive that is often built into mobile applications is based on earning rewards for behaving in a specified way. Often, those rewards are virtual: such is the case in the Ubigreen interface proposed by Froehlich et al. [2].

[2]For example, http://www.tfl.gov.uk/roadusers/cycling/14808.aspx
[3]See: http://www.freedompass.org

CONCLUSION

In this paper, we demonstrated how AFC records reveal hidden aspects of individuals' behaviours. To date, research on this data has focused on measuring the transport system's performance: in this work, we show how complimentary analysis can be used to measure the needs and choices of the system's end users.

We first addressed mis-alignments between self-reported and actual behaviour: we compared travel perceptions and smart card data from 85 London residents, and uncovered a number of differences that pave the way for future feedback-based travel applications. Many respondents were already using public transport more than they claimed to; we also determined that travel regularity is related to destinations and time of travel rather than the amount of travel. We also examined the extent that travellers (a) use modalities and (b) take trips from stations that they do not include in reports of their typical behaviour.

We then examined the extent that smart cards can be used to measure travellers' responses to various incentives, by comparing behavioural patterns with TfL's cost-related incentives. We showed that students do not tend to buy the discounted fares, and that switching to free travel does indeed encourage a higher use of the system. More generally, these results demonstrate that AFC data is a powerful too for measuring the success or failure of travel incentives.

Acknowledgements

The authors would like to thank Angela Sasse and Philip Inglesant for comments on the survey design. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-SST-2008-RTD-1) under Grant Agreement n. 234239.

REFERENCES

1. T. Choudhury et al. The Mobile Sensing Platform: An Embedded System for Capturing and Recognizing Human Activities. *IEEE Pervasive Computing*, 7(2), 2008.
2. J. Froehlich, S. Consolvo, T. Dillahunt, B. Harrison, P. Klasnja, J. Mankoff, and J. Landay. UbiGreen: Investigating a Mobile Tool for Tracking and Supporting Green Transportation Habits. In *ACM CHI*, Boston, USA, April 2009.
3. T. Campbell, E. Larson, G. Cogn, J. Froehlich, R. Alcaide, and S. Patel. WATTR: A Method for Self-Powered Wireless Sensing of Water Activities in the Home. In *ACM Ubicomp*, Copenhagen, Denmark, September 2010.
4. W. Jang. Travel Time and Transfer Analysis Using Transit Smart Card Data. *Transportation Research Board*, 2010.
5. H. Bryan and P. Blythe. Understanding Behaviour Through Smartcard Data Analysis. *Proceedings of the Institution of Civil Engineers: Transport*, 160(4), 2007.
6. L Frank, B Saelens, K. Powell, and J. Chapman. Stepping Towards Causation: Do Built Environments or Neighbourhoods and Travel Preferences Explain Physical Activity, Driving, and Obesity? *Social Science and Medicine*, 65, 2007.
7. N. Lathia, J. Froehlich, and L. Capra. Mining Public Transport Usage for Personalised Intelligent Transport Systems. In *IEEE ICDM*, Sydney, Australia, December 2010.
8. B. Ferris, K. Watkins, and A. Borning. OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit. In *ACM CHI*, Atlanta, GA, 2010.
9. J. Chan. Rail Transit OD Matrix Estimation and Journey Time Reliability Metrics Using Automated Fare Data. Master's thesis, MIT, Department of Civil and Environmental Engineering, June 2007.
10. K. Chu and R. Chapleau. Enriching Archived Smart Card Transaction Data for Transit Demand Modeling. *Journal of the Transportation Research Board*, (2063), 2008.
11. N. Lathia and L. Capra. Mining Mobility Data to Minimise Travellers' Spending on Public Transport. In *ACM KDD*, San Diego, California, August 2011.
12. S. Fujii and A. Taniguchi. Determinants of the Effectiveness of Travel Feedback Programs: A Review of Communicative Mobility Management Measures for Changing Travel Behaviour in Japan. *Transport Policy*, 13, September 2006.
13. K. Dziekan and K. Kottenhoff. Dynamic At-Stop Real-Time Information Displays for Public Transport: Effects on Customers. *Transportation Research Part A*, 41, July 2007.
14. D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft. Recommending Social Events from Mobile Phone Location Data. In *IEEE ICDM*, Sydney, Australia, December 2010.
15. K. Shampanier, N. Mazar, and D. Ariely. Zero as a Special Price: The True Value of Free Products. *Marketing Science*, 26(6):742–757, 2007.
16. S. Cairns, L. Sloman, C. Newson, J. Anable, A. Kirkbride, and P. Goodwin. Smarter Choices: Changing the Way We Travel. *Department for Transport, London*, July 2004.
17. D. Choujaa and N. Dulay. Predicting Human Behaviour from Selected Mobile Phone Data Points. In *ACM Ubicomp*, Copenhagen, Denmark, September 2010.
18. T. Sohn et al. Mobility Detection Using Everyday GSM Traces. In *ACM Ubicomp*, Orange County, USA, 2006.
19. Y. Zheng, Q. Li, Y. Chen, X. Xie, and W. Ma. Understanding Mobility Based on GPS Data. In *ACM Ubicomp*, Seoul, Korea, 2008.
20. C. Roth, S.M. Kang, M. Batty, and M. Barthelemy. Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. *PLoS ONE*, 6, 2010.
21. S. Azenkot, S. Prasain, A. Borning, E. Fortuna, R. Ladner, and J. Wobbrock. Enhancing Independence and Safety for Blind and Deaf-Blind Public Transit Riders. In *ACM CHI*, Vancouver, Canada, 2011.
22. T. Devinney, P. Auger, and G. M. Eckhardt. Values vs. value. http://bit.ly/eodkAT, March 2011.
23. E. Ampt. Understanding Voluntary Travel Behavior Change. *Australasian Transport Research Forum*, 22, 2003.
24. T. Dillahunt, G. Becker, J. Mankoff, and R. Kraut. Motivating Environmentally Sustainable Behavior Changes with a Virtual Polar Bear. In *Workshop on Pervasive Persuasive Technology and Environmental Sustainability (Pervasive)*, 2008.
25. J. Mankoff, D. Mattews, S. Fussel, and M. Johnson. Leveraging Social Networks to Motivate Individuals to Reduce their Ecological Footprints. In *HICSS*, Waikoloa, Hawaii, 2007.