DOES WORKING FROM HOME WORK?
EVIDENCE FROM A CHINESE EXPERIMENT

Nicholas Bloom[a], James Liang[b], John Roberts[c] and Zhichun Jenny Ying[d]

August 4th 2012

Abstract:
Over 10% of US employees now regularly work from home (WFH), but there is widespread skepticism over its impact highlighted by phrases like "shirking from home". We report the results of a WFH experiment at Ctrip, a 13,000 employee NASDAQ listed Chinese multinational. Call center employees who volunteered to WFH were randomly assigned to work from home or in the office for 9 months. Work from home led to a 13% performance increase, of which about 9.5% is from working more minutes per shift (fewer breaks and sick-days) and 3.5% from more calls per minute (attributed to a quieter working environment). Home workers also reported improved work satisfaction and their job attrition rate fell by 50%. After the experiment, the firm rolled the program out to all employees, letting them choose home or office working. Interestingly, only half of the volunteer group decided to work at home, with the other half changing their minds in favor of office working. After employees were allowed to choose where to work, the performance impact of WFH more than doubled, highlighting the benefits of choice alongside modern management practices like home working.

Keywords: working from home, organization, productivity, field experiment, and China

Acknowledgements: We wish to thank the management of Ctrip, as well as Jennifer Cao, Mimi Qi, Maria Sun and Edison Yu for their help in this research project. We thank Chris Palauni for organizing our trip to Jet Blue, and David Butler, Jared Fletcher and Michelle Rowan for their time discussing the call-center and home-working industry. We thank in particular our discussants Mushfiq Mobarak, Rachael Heath and Sabrina Pabilonia, and seminar audiences at the AEA, Brown, the London School of Economics, Stanford GSB and the World Bank for comments.

Conflict of interest statement: We wish to thank: our discussants Mushiq Mobarak and Rachel Heath; and Stanford Economics Department, Stanford GSB, and the Toulouse Network for Information Technology (which is supported by Microsoft) for funding for this project. No funding was received from CTrip. James Liang is the co-founder, ex-CEO and current Chairman of Ctrip. No other co-author has any financial relationship (or received any funding) from CTrip. The results or paper were not pre-screened by anyone.

[a] Stanford Economics, SIEPR, CEPR and NBER; [b] Ctrip, [c] Stanford GSB and SIEPR, [d] Stanford Economics

I. INTRODUCTION

Working from home (WFH) is becoming an increasingly common practice. In the United States, over 10% of the workforce reports working from home at least one day a week, while the proportion primarily WFH has almost doubled, from 2.3% in 1980 to 4.3% in 2010 (Figure 1a). At the same time, the wage discount (after controlling for observables) from working exclusively at home has fallen, from 30% in 1980 to zero in 2000 as WFH moved from being predominantly in only low-skilled jobs to encompass a wider set of occupations (Oettinger, 2010). Home-based workers now span a wide spectrum of occupations, ranging from sales assistants to managers and software engineers, with a correspondingly wide range of incomes (Figure 1b).

The balance between work and private home life has also received increasing attention as the number of households in the US with all parents working has increased from 25% in 1968 to 48% in 2008 (Council of Economic Advisors, 2010). These rising work pressures are leading governments in the US and Europe to investigate ways to promote work-life balance. For example, the Council of Economic Advisers (CEA) published a report launched by Michelle and Barak Obama at the White House in the summer of 2010 on policies to improve work-life balance. One of the key conclusions in the executive summary concerned the need for research to identify the trade-offs in work-life balance policies, stating:

> "*A factor hindering a deeper understanding of the benefits and costs of flexibility is a lack of data on the prevalence of workplace flexibility and arrangements, and more research is needed on the mechanisms through which flexibility influences workers' job satisfaction and firms' profits to help policy makers and managers alike*" (CEA, 2010)

Not surprisingly, given this lack of research, many firms are uncertain about whether to permit working from home. As a result, firms in the same industry have adopted different practices. For example, in the U.S. airline industry, JetBlue allows all regular call-center employees to work from home, while Delta and Southwest allow no home working and United has a mix of practices. More generally, Bloom, Kretschmer and Van Reenen (2010) report 30% of US and 33% of European manufacturing firms offer opportunities for at least some managers to work from home, with wide variation in adoption rates within every 3-digit SIC code surveyed. They find similar variation in the adoption of other practices affecting work-life balance practices like job-sharing, part-time working, flexi-time and extended maternity leave within every industry, with no consensus around what defines a "best-practice".

CTrip International Corporation ("Ctrip") – China's largest travel agency with 13,000 employees and a $5bn valuation on NASDAQ – was interested in allowing its call-center employees to work from home. The perceived benefits included reducing office rental costs, which were becoming increasingly onerous due to rising rental rates at the Shanghai headquarter, reducing their 50% annual rate of attrition among call-center workers and gaining access to potential employees who lived too far from the Ctrip office to commute to work there. The executives' concern was that allowing employees to work at home, away from the supervision of their shift managers, would have a negative impact on their performance.

Given the uncertainty surrounding the effects of working from home, the firms' leaders decided to run a randomized controlled experiment. We assisted in designing the experiment and have had complete access to the resulting data and to data from surveys conducted by the firm. We have also conducted various surveys ourselves and numerous interviews with employees and managers.

In summary, Ctrip decided to run a nine-month experiment on working from home. They asked the 996 employees in the airfare and hotel departments of the Shanghai call center whether they would be interested in working from home four days a week. Approximately half of the employees (508) were interested. Of these, 255 were qualified to take part in the experiment by virtue of having at least six months tenure, broadband access and a private room at home (in which they could work). After a lottery draw, those with even birthdays were selected to work at home while those with odd birthdates stayed in the office to act as the control group.

Throughout, the workers were organized in "teams",[1] each under a team leader, and all members of the team worked the same shifts. Assignments to teams remained unchanged through the experiment, so some members of a given team would be in the treatment group and others in the control. The home and office employees in each team had to work the same shift because they worked under a common team manager. The two groups also used the same IT equipment, faced the same work order flow from a common central server, and were compensated under the same pay system. Hence, the only difference between the two groups was the location of work.[2] This allows us to isolate the impact of working-from-home (flexi-place) versus other practices that are commonly bundled alongside this practice, such as flexible work hours (flexi-time).

We found four main results. First, the performance of the home workers went up dramatically, increasing by 13% over the nine months of the experiment. This improvement came mainly from a 9.5% increase in the number of minutes they worked during their shifts (i.e., the time they were logged in to take calls). This was due to a reduction in breaks and sick-days taken by the home workers. The remaining 3.5% improvement came from home workers increasing calls per minute worked. In interviews, the workers attributed this gain to the quieter working conditions at home. Second, there were no spillovers to the rest of the group – interestingly, those remaining in the office had no drop in performance despite losing the treatment lottery. Third, attrition fell sharply among the home workers, dropping by 50% versus the control group. Home workers also reported substantially higher work satisfaction and had more positive attitudinal survey outcomes. Finally, at the end of the experiment, the firm estimated it would have saved about $2,000 per year per employee working at home, leading it to offer the option to work from home to the entire firm. This allowed the treatment and control groups to re-select their working arrangements. Almost half of the treatment group changed their minds and returned to the office, while two thirds of the control group (who initially had all volunteered to work from home) decided to stay in the office. This selection led to much larger long-run impacts from working at

[1] There is no sense in which the call-center jobs involved "teamwork" and there was no group-based pay, but we stick with the term "team" because that was what Ctrip called the work groups.

[2] This of course had implications that were potentially relevant to the experiment. In particular, employees at home did not have on-going, immediate contact with the team leaders and they worked in a different environment than those in the office. We discuss these points more below.

home, as workers with relatively better performance at home remained at home while those performing relatively poorly at home returned to the office.

This experiment thus highlights the extensive learning by both the firm and employees around the adoption of a modern management practice like working from home. *Ex-ante*, both groups were unsure about its impact, and the 9-month experiment and subsequent roll-out process were essential for their ability to evaluate it. These gradual learning effects are one factor behind the slow adoption of modern management practices, and we see the results as similar to the adoption process for other types of innovations, like hybrid seed-corn as emphasized in Griliches' (1957) classic article.

This experiment is also unique as the first randomized experiment on working from home. As such, it provides much more solid evidence than what has been available from the previous case-study and survey research on the subject. A second way in which this study is unusual is the fact that it is a randomized controlled experiment within a large firm. In running this we were granted exceptional access not only to data but also to the Ctrip management's thinking about the experiment and its results. This was because one of the co-authors of this paper--James Liang--co-founder, first CEO and current Chairman of Ctrip, was also a doctoral student at Stanford at the time. As a result, this paper benefits from unusually detailed insight into the adoption of a new management practice in a large, multinational firm.

The paper connects to three strands of literature. First, there is a strand of literature on the adoption of work-life balance practices, which is based primarily on case-studies and surveys across firms. These tend to show large positive associations of adoption with lower employee turnover and absenteeism, and with higher productivity and profitability (for example, see the surveys in CEA 2010, Bloom, Kretschmer and Van Reenen 2010, Bloom and Van Reenen 2011, and Oyer and Lazear 2012). But these studies are hard to evaluate because of the non-randomized nature of these programs. This is both true in terms of the selection of firms into working-from-home programs, and also the selection of employees to work at home. For example, as we show in Table 7 when CTrip allowed a general roll-out of home-working, we see high-performing employees choosing to move home and low-performing employees choosing to return to the office, so that the full impact of working from home, including selection effects, looks twice as large as the simple experimental impact.[3]

More generally, there is a long literature on the puzzling dispersion of productivity between firms (see the literature from Walker 1887 to Leibenstein 1966 to Syversson 2011 and Gibbons and Henderson 2012). This paper provides one rationale for this dispersion, which is the slow spread of modern management practices, including those addressing work-life balance, like working from home. The adoption of practices aimed at improving work-life balance is highly variable across firms in the US and Europe because of the uncertainty about their impact, but they have potentially large effects on measured productivity. For example, based on the methodology that is usually used to measure productivity using Census data, CTrip would have increased productivity by 30% after introducing working from home, even before accounting for selection effects.

[3] Strikingly, this is the same ratio of benefits from selection to total benefits from the intervention found in Lazear's famous study of Safelite Auto Glass (Lazear 2000).

Finally, there is also the connection to the urban economics literature. Firstly, reducing the frequency of commuting will reduce vehicle miles travelled, lowering emissions, but also population centrality as people move out into the suburbs (Bento, Cropper, Mobarak and Vinha, 2005). For example, Jet Blue allows home-based call center employees to be up to 3 hours drive from the office since they only need to come into the office one day per month, so that many of them now live in rural areas outside Salt Lake City (where the firm is headquartered). Secondly, working from home is part of wider impact of IT on firm-fragmentation arising from the increasing ease of long-distance communicating (Rossi-Hansberg, Sarte and Owens, 2099). For example, CTrip is now setting up regional offices to employ workers in lower-wage inland Chinese cities using the same working-from home technology they used in this experiment.

Section II describes the experiment in more detail, while section III presents the firm results, and section IV the impact on employees, while section V discussed the roll-out and finally section VI provides a set of concluding comments.

II. THE EXPERIMENT

II.A. The Company

Our experiment took place at Ctrip International Corporation, a leading travel agency in China with operations also in Hong Kong and Taiwan. Ctrip aggregates information on hotels and flights, makes reservations and obtains tickets for clients, and generates revenue through commissions from hotels, airlines and tour operators. The services provided by Ctrip are comparable to Expedia, Orbitz or Travelocity, although, because of lower Internet penetration in China, Ctrip did much more of its business on the telephone. Ctrip was established in 1999 and was quoted on NASDQ in 2003, and was worth about $5bn at the time of the experiment. It was the largest travel agent in China in terms of hotel nights and airline tickets booked, with over 50% market share in 2010. Exhibit A displays photos of the Ctrip Shanghai office, a modern multi-story building that housed the call center in which the experiment took place, as well as several other CTrip divisions and its top management team. The firm also operated a second, larger call center in Nan Tong, outside Shanghai, which employed about 2/3rds of the 7,500 call center staff. Both locations operated in the same fashion under the same procedures.

Call center employees were organized into small teams of around 10 to 15 people (mean of 11.7 and median of 11), grouped by department and the type of work. There were four jobs in each of the two departments – hotel and airline – involved in the experiment. These were order takers, who answered customer calls and took orders, entering them into the Ctrip information system; order placers, who dealt with the airlines and hotels then notified the clients; order correctors, who resolved problems such as when a flight was cancelled, plus a night shift that both placed and corrected orders. The members of a given team sat together in one area of the floor, typically occupying an entire aisle. Each team member worked in a cubical with equipment including a computer, a telephone and a headset. When team members were ready to start work, they logged on to CTrip's IT system and, in the case of order takers, client calls were automatically routed into their headsets. Order placers and order correctors also were allocated tasks automatically. The allocations between the two Shanghai and Nan Tong call centers were handled centrally, using a standard call queuing system. When employees wanted to take a break, they logged out of the system. The team leaders patrolled the aisles to monitor employees' performance as well as help resolve issues with reservations, provide ongoing training and give emotional support to employees dealing with difficult clients.

The employees typically worked 5 shifts a week, scheduled by the firm ahead of time. All members of a team worked on the same schedule, so individuals did not choose their shifts. The firm adjusted the length of the shifts depending on the anticipated volume of the bookings.

Monthly salary was composed of a flat wage and a bonus portion. The flat wage depended on seniority, education and prior experience, and averaged around ¥1300 per month. The bonus portion mainly depended on monthly performance, and averaged about ¥1300 per month. The bonus was fundamentally a linear function of call and order volumes, but with small adjustments for call quality (penalties were applied for call quality scores below certain thresholds) and shift type (night shifts, for example, were paid a higher flat rate). Promotion to team-leader was also based on performance, so both salary and career concerns provide incentives for employees to perform well.

Ctrip was interested in running the experiment to investigate the impact of allowing employees to work from home. They believed allowing employees to work from home would save office space, reduce turnover (saving on recruiting and training costs), and reduce labor costs by tapping into a wider pool of workers, such as people living too far outside Shanghai to commute in on a daily basis. But the leaders of the firm were worried about the impact on performance of allowing employees to work from home. Most of the call center workforce was made up of younger employees, many of whom might have struggled to remain focused working from home.

Since no other Chinese firm had successfully moved to allowing home-working among its call center employees, there was no local precedent. In the US, the decision to allow employees in call centers to work from home varied across firms, even those within the same industry, suggesting a lack of any consensus on its impact. Meanwhile, the prior academic literature on call centers also offered limited guidance, being based on case studies of individual, firm-level interventions. So management decided to run an experiment.

II.B. The Experimental Design

The experiment took place in the airfare and hotel booking departments in the Shanghai call center. The treatment in our experiment was to work 4 shifts a week at home and to work the 5th shift in the office on a fixed day of the week determined by the firm. Treatment employees still worked on the same schedule as their teammates because they had to work under the supervision of the team leader (who is always office-based), but they operated from home for 4 of their 5 shifts. For example, in a team the treatment employees might work from home from 9am to 5pm on Monday, Tuesday, Wednesday and Friday and in the office from 9am to 5pm on Thursday. The control employees from that team would work in the office from 9am to 5pm on all five days. Hence, the experiment only changed only the location of work, not the type or the hours of work. Because all incoming phone-calls and work orders are distributed by central servers, the work flow was also identical between work and home locations. Home workers also used the same, Ctrip-provided computer terminals, communications equipment and software, faced the same pay structure and promotion procedures and undertook the same training as the control group (although for the treatment employees this occurred only the day they were in the office).

Importantly, individual employees are not allowed to work overtime outside their team shift as it would require their team leader to supervise their work. Hence, entire teams could have their hours changed – for example all teams had their shifts increased during the week before Chinese New Year – but individuals were not able to work overtime on their own. Thus, eliminating commuting time, which was 80 minutes a day for the average employee, did not permit the treatment group to work overtime and so this is not a factor directly driving the results.

Three factors other than location did differ (unavoidably) between treatment and control. First, the treatment group's spending less time commuting meant that they would sometimes be able to take care of personal and family responsibilities without taking breaks or leaving early from work. As we will see, this appears to have had a significant effect. Second, the treatment workers did not have as much support from their team leaders, because technological limitations meant that they could not simultaneously talk to their supervisors and deal with clients online. If anything, this presumably reduced the effectiveness of the treatment workers and strengthens the results. Finally, the work environment differed between treatment and control. The former were working alone, typically in a quieter environment. This had some negative effects on willingness to work from home, but positive effects on productivity.

In early November 2010, employees in the airfare and hotel booking departments were informed of the working from home program. They all took an extensive survey on demographics, working conditions and their willingness to join the program. When inquired of their willingness to join the program, employees were not told the set of criterion that they would have to qualify in order to participate in the program. Employees who were both willing and qualified to join the program were recruited for the experiment. Of the 996 employees in the airfare and hotel booking departments, 508 (51%) volunteered for the experiment, with those with a more expensive and longer commute, with less tenure in the firm, with less education and with their own bedroom significantly more likely to want to work from home (see Table 1). Importantly, prior-performance (measured by the gross-wage given that almost 50% of salary is performance related pay) was not predictive for the take-up of working from home. This helped to assuage one concern of the firm that lower performing employees would be more tempted to work from home to avoid the direct supervision of their team-leaders.

Interestingly 49% of employees did not opt to work from home despite the considerable saving in commuting time and cost. The major reason given for this in later interviews was the loneliness of working from home and the lack of opportunities to socialize in the office and after work.

To qualify, an employee also needed to have tenure of at least 6 months, have broadband Internet at home to connect to the network, and an independent workspace at home during their shift (such as their own bedroom). Among the volunteers, 255 (50%) of the employees met the eligibility requirements and were recruited into the experiment.

The treatment and control groups were then determined from this group of 255 employees through a public lottery. Employees with an even birth date (a day ending 2, 4, 6, 8, etc.) were selected into the treatment and those with an odd birth date were in the control group. This selection of even birthdates into the treatment group was randomly determined by the Chairman, James Liang, by drawing a ping pong ball from an urn in a public ceremony one week prior to the experiment's start date (see Exhibit B).[4] Qualified employees with even birth dates who had chosen to be in the experiment group were notified and equipment is installed at each treatment participant's home the following week. Qualified employees with odd birth dates who had chosen to be in the experiment became the control group. The experiment commenced on December 6, 2010.

The experiment lasted for 9 months, and all treatment employees had to remain at home for this period, even if they changed their mind and wanted to return to the office. On August 15, 2011, employees were notified that the experiment had ended and Ctrip would roll out the experiment to those who were qualified and wanted to work at home in the airfare and hotels departments on September 1st, 2011.

Throughout the experiment, employees were told the experiment would be evaluated to guide future company policies, but they did not learn the actual policy roll-out decision until August 15th. Because of the large scale of the experiment and the lack of dissemination of experimental results beyond the management team, prior to the roll-out decision, employees were uncertain about what that decision would be. Employees in the treatment group who wished to come back to work in the office full-time were only allowed to do so after August 15th, while control workers had to stay in the office for the full duration of the experiment. Hence, the treatment and control assignments were fixed for the full 9 months.

Figure 2 shows compliance with the experiment throughout the experimental period, and after the general roll-out through May 2012. During the experiment, the percentage of treatment group working at home hovered between 80% and 90%. The compliance did not reach 100% because in a few cases the broadband speed was not fast enough to support working from home, but more often because employees moved apartments and lost access to their own room[5]. Since compliance was not perfect, our estimators take even birth date status as the treatment status, so we estimate an intention to treat result. Given we are interested in evaluating the impact of a policy of allowing home-working, this seems appropriate.

After the experiment, we see about 50% the treatment group immediately decided to return to the office. They did this despite having to incur the financial and time costs of commuting, with the main reason given for this in interviews being the loneliness of working from home. Strikingly, only about 35% of the control employees – who also all initially volunteered to work from home – actually move home when they were allowed to do so. Again, the main reason they gave for

[4] It was important to have this draw in an open ceremony so that managers and employees could not complain of "favoritism" in the randomization process. The choice of odd/even birthdate was made to ensure the randomization was straightforward and transparent.

[5] In all estimations, we use the even birthdate as the indicator for working-at-home so these individuals are treated as home workers. In a probit for actually working from home during the experiment, none of the observables are significant, suggesting that returning to the office was effectively random. One reason is that the IT group policed this heavily to prevent employees fabricating stories to enable them to return to the office.

changing their mind was concerns over being lonely at home. Finally, we also see that about 10% of the workers who did not initially volunteer changed their minds after the experiment and decided to work from home.

It is worth noting that the firm's management was surprised by two things in these numbers. First, they were struck by how many employees changed their minds about working from home. About 50% of the volunteer group and 10% of the non-volunteer group switched preferences after the 9-month experiment. Employees reported that it after working from home for a few months they started to get lonely and wanted to return to the office.

Second, despite the time and financial savings from not having to commute, more than half of the workers eligible to work at home decided to work in the office, suggesting they place a high value on social interactions at work (Hamermesh, 1990). This is particularly striking because, as we note below, we find no negative impact of home working on any other outcomes like call quality or promotions.

II.C. Data Collection

Ctrip had an extremely comprehensive central data collection system, as its founding team came from Oracle with extensive database software experience. The majority of data we used in our paper were directly extracted from the firms' central database, providing extremely high data accuracy. The data we collected can be categorized in 7 fields: performance, labor supply, attrition, promotions, reported employee work satisfaction, detailed demographic information, and survey information on attitudes towards the program.

Performance measures vary by the broad type of workers – the 137 order takers and the 118 order placers, order correctors and night shift workers (details in Appendix 1). Order takers took incoming calls. Their key performance measures were the number of phone calls answered and number of orders taken. The other three groups placed calls, and their key measures were the numbers of different types of calls made. For order takers, we can also accurately measure time spent working in terms of minutes on the phone because we have logs of phone calls and call lengths from the central database of Ctrip. The firm used these measures to monitor the work of their employees. We also calculated phone calls answered per minute on the phone as a measure of labor productivity for these workers.

We have daily key performance measures of all employees in the airfare and hotel booking departments from January 1st, 2010 onwards. We also have daily minutes on the phone for order takers during the same period. We have daily records of hours of leave for the airfare department, and the date and reason of employees in the experiment quitting the firm. The firm also ran internal surveys of the employees during the experiment on work exhaustion and, positive and negative attitudes (see details in Appendix A2). Finally, we conducted two rounds of surveys, in November 2010 and August 2011, to collect detailed information on all the employees in the two departments including basic demographics, income, and attitudes toward the program.

III. IMPACT ON THE FIRM

We analyzed the effect of home-working both in terms of its impact on the firm, which we cover in this section, and the impact on the employees, which we cover in the next section.

III.A. Individual Employee Performance

We started by estimating the intention to treat effect on employee performance via equation (1)

$$Employee\ Performance_{i,t} = \alpha Treat \times Experiment + \beta_t + \gamma_i + \epsilon_{i,t} \quad (1)$$

where *Treat* is a dummy variable that equals 1 if an individual belongs to the treatment group defined by having an even-numbered birthday; *Experiment* is a dummy variable that equals 1 for the experimental period December 6th to August 15th; and *Employee Performance* is one of the key measures of work performance, including an overall performance z-score measure, log of weekly phone calls answered, log of phone calls answered per minute on the phone, and log of weekly sum of minutes on the phone. Finally, β_t reflects a series of week dummies to account for seasonal variation in travel demand, such as the World Expo in 2010 and the Chinese New Year, and γ_i reflect a full set of individual fixed effects.

To make performance of different types of workers comparable, we use performance z-scores. For each individual we subtract the pre-experiment mean for their worker type, and divide by the pre-experiment standard deviation for their worker type. Hence, this normalized z-score measure has a mean 0 and standard deviation 1 across all employees within each type of worker during the pre-experiment period.

In column (1) of Table 2, overall performance of the treatment group is found to be 0.226 standard deviations higher than the control group after the experiment started, significant at the 1% level. The largest group of workers we have in our sample are the 137 order takers. If we limit the sample to them, we can use phone calls answered as the key performance measure for all the order takers. The z-scores of phone calls account for different volume and average length of phone calls in two departments. In column (2), we look at just the phone calls performance measure and find this is 0.263 standard-deviations better for the treatment group. In column (3), we look at the log of phone calls and find these are 0.122 higher, so that treatment employees were making 13% (noting that 13%=exp(0.122)) more phone calls.

We can also see these results in Figure 3a where we plot the raw number of phone calls per week for the treatment and control groups from Jan 1st 2010 until the end of the experiment in August 15th 2011. Before the experiment started, the treatment group trends closely together with the control group, both of which bounce around due to seasonal fluctuations in demand. But once the experiment began, the treatment group started to outperform the control group, answering about 40 more phonecalls per person per week. Figure 3b plots the cross-sectional distribution of performance for treatment and control groups at 3 months, displaying a broad distributional improvement from working-from-home (rather than the results being driven by a few outliers).

We further decomposed the difference in performance observed in column (3) into phone calls answered per minute on the phone (a measure of productivity), and minutes on the phone (a measure of high-frequency labor supply). In column (4), we found treatment employees were making 3.4% (note that 3.4%=exp(0.033)) more phone calls per minute, which the employees working from home identified as resulting primarily from home being quieter than the office. They told us this meant it was easier to hear the customers, so they did not have to ask them to repeat themselves as often and could process the information more quickly. This suggestion matches the psychology literature which has shown that background office noise can reduce cognitive performance (see, for example, Banbury and Berry, 1998).

But, the biggest factor increasing the home-workers performance is that, as shown in column (5), they worked 9.4% (9.4%=exp(0.089)) more minutes per day. This was despite the fact that home and office workers both worked the same nominal shift – for example, 9am to 5pm on Monday to Friday – as members of both groups worked in the same team under the same team manager. The reason home-workers could increase minutes on the phone was within their shift they were logged on and available to take calls for more time, meaning they were taking less time-off during their shift.

III.B. Individual Employee Labor Supply

In Table 3, we investigate the factors driving this increase in minutes worked within each shift. Because we have accurate records of hours of leave from the airfare booking department only, we limit the sample further to the 89 order takers in the airfare department. Column (1) repeats the results from the final column of Table 2, while Column (2) of Table 3 shows that these order takers show a very similar increase to the full group.

Columns (3)-(5) break this difference in minutes on the phone down into three pieces. In column (3), we look at whether treatment workers spent more minutes on the phone per hour logged in,[6] column (4) looks at whether they were logged in for more hours per day worked, and column (5) looks at whether they worked for more days.

What we found is that in column (3), there is no difference between the number of minutes on the phone while logged-in for the treatment and control employees. This is not surprising because both groups operated using the same call routing server and on the same queuing system.[7]

Column (4) shows that about two-thirds of the difference in the time on the phone was accounted for by home-workers logging in for more hours per day worked. This is because: (a) they started work more punctually and left early less often. They attributed much of this to their avoiding the effects of events that disrupted commuting like the heavy snow in Shanghai in February 2011; and (b) they took shorter lunch breaks because they were usually eating on their own rather than with colleagues in the canteen. They also were able to fit in personal matters like doctor's appointments without leaving early. Finally, in column (5) we see that the other third of the

[6] Note that sometimes employees would not be taking calls when they are logged in if demand is low, so that time logged in and time on the phone are not necessarily the same (the former is higher when demand is low).

[7] Moreover, it shows that home-workers are not picking busier times to log in to the system (i.e. they are not timing their breaks to coincide with quiet periods when demand is lower). I thank Wouter Dessein for pointing this out.

difference in time worked between treatment and control was because treatment employees worked more days because they took fewer sick-days. Employees explained this was because they continued to work at home when they felt somewhat ill but would not have felt up to commuting into work.

III.C. Quality, Spillovers and Potential Hawthorne Effects

One question is whether quality of the service was compromised for the increase in output in the treatment group. We constructed two quality measures: conversion rates and weekly recording scores. Conversion rates were calculated as the percentage of phone calls answered resulting in orders, while the weekly recording scores came from the 1% of phone-calls that are randomly evaluated by an external monitoring team. In summary (with the full details in table A3 in the appendix), we find no impact of working from home on call quality using either measure.

Another related question is whether the improvement associated with working from home came from an improvement in the treatment group or from a deterioration in the control group. Perhaps the gap between treatment and control was caused not by the treatment group performing better but by the control group performing worse after they "lost" the randomization lottery. The group winning the treatment lottery saved themselves 9 months of commuting time and costs, a substantial gain worth about 17% of their salary, evaluated at their CTrip wage rate.[8]

We collected data on two other "quasi" control groups to answer this question. The first group is the eligible employees in the Nan Tong call center. This was CTrip's other large call center, located in Nan Tong, a city about 1 hour drive outside of Shanghai. This call center also had airfare and hotel departments, and calls were allocated across the Shanghai and Nan-Tong call centers randomly from the same central server. The second group was the 253 eligible employees who did not volunteer to participate in the WFH experiment in the Shanghai call center. These were the individuals that were eligible for the experiment (own room, 6+ months of tenure and broadband), but did want to work from home. We think these two groups are comparable to the treatment and control groups for two reasons. First, all four groups face the same demand for their service. Second, they all meet the requirements for eligibility to participate in the experiment. Figure 4 shows the performance of the eligible group in the Nan Tong call against the treatment and control groups, highlighting how they all tracked each other well before the experiment. After the experiment started, the performance of the Nan Tong group was similar to that of the control group.

More formal comparisons of these alternative control groups are also reported in Table 4. Results in the top panel of Table 4 compare the treatment and control groups to Nan Tong, showing differences in overall performance, efficiency and labor supply with the control group were statistically insignificant from zero. The bottom panel compares treatment and control group to the eligible non-experimental group in Shanghai. Again, we found no difference between the control group and the eligible non-experimental group. These results suggest that the gap between the treatment and control group reflects an improvement in the performance of the treatment group rather than any deterioration of the control group. That is, although the control

[8] The average employee makes about $100 per week for a 40 hour week. The commuting time is 40 minutes each way and the out of pocket cost $0.5 on average. Hence, the saving in time is about $13 a week in time costs and about $4 per week in out of pocket costs.

group and the treatment group work in the same team, we find – perhaps surprisingly – no evidence of the control group's being discouraged by losing the working-from-home lottery.

We also looked for spillovers by examining the variation in the number of individuals randomly assigned to treatment across the groups within the Shanghai office. Because groups are small, random variations in the number of employees with even and odd birthdays generated variations in the fraction of employees in a team who got to work at home. We used this (the share of evens in the eligible volunteered group within each team) to instrument for the share of all employees working from home, and investigated the impact of this on the team's performance. As we show in Table A4, we again found no evidence for spillovers across individuals from home-working.

Finally, another explanation for the superior performance of the treatment group are Hawthorne effects (they were motivated by the experiment), possibly deliberately so that the firm would roll out WFH permanently. We should note three things, however, that make this appear unlikely. First, there were 122 employees working from home, so each individual employee has little impact on the evaluation of the experiment. Second, the home-based employees performed even better after the experiment ended. Finally, the firm was itself so convinced that the success of the experiment was not due to Hawthorne type effects that it rolled out WFH across both divisions.

III.D. Post-Experiment Selection

In August 2011, the management estimated that each working from home employee saved CTrip about $2,000, so they decided to immediately roll out the option to work from home to the entire hotel and airfare departments. Employees in these departments were notified that the experiment had ended and they were entitled to choose their location of work – control employees who still wished to could move home, and treatment employees that wanted to return to the office could do so.

As shown in Figure 5 – which plots the difference in normalized phone-calls between home and office workers – post-experiment selection substantially increased the performance increase from working from home. The differential increase in phone calls (versus the pre-experiment baseline) from home-working was about 0.2 standard-deviations during the experiment, rising to about 1 standard deviation within 6 months after the experiment. This is also evaluated in Table 5 which estimates the performance impact of working-from home during and after the experiment. As we see in columns (1) and (2) after the experiment the average impact rises from about 0.148 to 0.273, with column (3) showing this impact appears to be increasing over time as indicated by Figure 5. Finally, column (4) reports similar results for a balanced panel of employees (dropping anybody that quits before the end of May 2012), showing that it is sorting of employees between home and the office rather than differential attrition that is driving the approximate doubling of the impact of working-from home during the experimental roll-out period.

This sorting is driven by treatment workers who had performed relatively badly at home returning to the office. This is shown in Table 6, columns (1) to (4), which run probits on whether a treatment worker returns to the office. The results show that treatment workers who performed relatively worse at home versus the office returned to the office. This was despite the fact that all treatment workers had initially volunteered to work from home, suggesting that many of them subsequently discovered home working was not as attractive as they initially believed (nothing they receive performance pay).

IV. IMPACT ON THE EMPLOYEES

III.A. Employee' self-reported outcomes

Ctrip management was also interested in how employee self-reported wellbeing was affected by the program. They thus ran two sets of surveys: the satisfaction survey and the emotion survey. Details of survey questions and methodology are listed in Appendix A2, but in summary these were standard employee satisfaction tests developed by Christina Maslach and Susan Jackson in the 1970s (see for example Maslach and Jackson, 1981). The satisfaction survey was conducted five times throughout the experimental period: once in early November before the randomization took place and four times after the experiment had started. Because the employees were unaware of the assignment at the initial survey date, the first survey was a credible baseline. The first three columns of Table 7 show three different satisfaction measures. The treatment group reported no difference in satisfaction levels from the control group at the first survey, but the treatment group reported statistically significantly higher satisfaction levels throughout the experiment.

The emotion survey was conducted every week. The first week was conducted in late November 2010, before the experiment began but after the randomization, so that individuals had been informed of their status in the treatment or control groups. Interestingly, the treatment group already reports higher positive attitude (significant at the 10% level), less negative attitude and less exhaustion from work. This group had yet to move home, so this difference is entirely due to the control group's learning they lost the randomization while the treatment group learned they had won, and highlights the importance of comparing our treatment groups with other controls groups like Nan-tong and the non-volunteer group. After starting the experiment, the gap between the treatment and control group rose further, so that the treatment group reported statistically significantly higher positive attitude and less work exhaustion. Of course, their total work plus commute time was lower on average than the control group.

IV.B. Attrition

One of the key initial reasons Ctrip was interested in running the experiment was to see if working from home would help retain workers. The turnover rate among Ctrip call center representatives had historically hovered around 50% per year, which was typical of the call center industry in China[9]. Management estimated that hiring and training a call center representative cost on average $2000, about 6 months' salary of an average employee. Figure 6 plots the cumulative attrition rate of treatment and control group separately over the experimental period. Shortly after the commencement of the experiment, cumulative attrition rates diverged between the two groups and the difference is statistically significant. By the end of the experiment, the total attrition rate in the treatment group (17%) was less than half of that in the control group (35%).

[9] 2010 Report on Chinese Call Center Operation and Management. Note that CTrip could in principle fire employees, but this was rare and no employees in these two divisions were fired over this period as far as we are aware.

We further tested whether selective attrition existed by running probit regressions in Table 8. The dependent variable is whether an employee quit the job during the experimental period between December 6^{th} 2010 and August 15th 2011. Column (1) confirms the finding in Figure 6, that treatment employees rate of attrition was about half that of the control group. In column (2), we tested whether employees with worse performance were more likely to leave the firm from the treatment group compared to the control group, but we found no supporting evidence. Not surprisingly, we did find, however, that younger employees and those with higher commuting costs were more likely to quit.

In column (3), we used the same specifications as in column (2), but replaced the pre-experiment performance with experiment performance During the experiment. This is the average of individual weekly performance z-scores during the experimental period from December 6^{th} 2010 to August 15th 2011. We found that low performers were significantly more likely to quit, particularly those in the control group. In columns (4) and (5), we estimated the impact of experimental period performance on quitting in the treatment and control groups separately and found a significant impact only for the control group. From interviewing the employees, we heard that control group employees who underperformed tended to quit for other call-center that they believed would pay better. Treatment employees, however, were much less likely to quit because no other home-working jobs existed, substantially reducing selection from the treatment group.

This differential attrition, of course, also raises the question of whether our estimated impact of WFH is biased. To address this issue, we use the Lee (2010) bounds estimator. This provides upper and lower bounds on the differential selection on performance across groups, assuming that selection into the control group monotonically increases attrition. This allows us to generate two bounds – the upper bound that assumes that the extra attrition in the control group is based on a negative correlation between performance (as we saw in Table 7) while the lower bound assumes a positive correlation (the reverse of what we see in Table 7, but included for completeness). We see that the upper bound lies above the actual treatment-control estimated impact, suggesting that the actual treatment effect on attrition is, if anything, larger than we estimated, because the attrition of the worst performers from the control group biases our results down.

IV.C. Promotions

One possible negative effect from working at home is that long-run career performance could be damaged by less on-the-job assistance and training from team leaders and less “face-time” in the office, making it harder for home-based workers to achieve a promotion. To investigate this, we collected promotions data on the 255-employee experimental sample. In summary, during the period from the start of the experiment in December 2010 until May 2012, a total of 8 employees from the treatment group received promotions and 6 from the control group. Neither this raw difference nor the coefficient on treatment in promotion probits including or excluding demographic controls was significant. Thus, at least over the period of 18 months from the beginning of the experiment until May 2012, we found no negative impact of working at home 4 days a week on employees’ ability to get promoted.

V. PROFIT, PRODUCTIVITY AND FIRM LEARNING

One of the most interesting aspects of the experiment was the learning process for both the firm and the individual employees on the costs and benefits of working from home. Both groups were initially unsure about its impact, because a practice given this had never previously been adopted by other Chinese travel agents or call centers had never offered this option. However, we were able to monitor both management's and employees' learning over the course of the experiment because of our extensive access to the CTrip's management team and frequent employee surveys and interviews. Before discussing this we first present the estimated impacts on firm profits and productivity from allowing employees working-from home.

V.A. Profit and productivity impact

The firm saw working from home as a way to save on office costs, but was worried that employees would shirk at home or that call quality would decline due to multi-tasking on other activities which are prohibited in the office like playing computer games or watching TV. Running the experiment revealed, however, that working from home actually generated an improvement in employee performance, worth about $375 per employee per annum (evaluated at the 13% performance improvement from the Table 3). In addition, they estimated office cost savings of about $1250 per employee and reduced turnover savings of about $400 per employee per annum. Hence, given the saving of about $2000 per employee, the firm rolled the program out in August 2011, accompanied by an aggressive poster campaign to persuade employees to take up the home working option.

A related question is what was the impact on total factor productivity (TFP)? We estimate TFP would increase by about 30% from moving every employee home, using the methodology adopted on US Census data by, for example, papers like Foster, Haltiwanger and Syversson (2008) and Syversson (2011).

This 30% rise in measured TFP comes from three sources. First, output (as measured by the number phonecalls) increased 13% from working from home. While 9.5% of this increase comes from employees working more hours, this increase in attendance would not be measured in US Census survey data, since this collects information on shift-hours (i.e. 40 hours per week), not actual hours worked. Second, the reduction in attrition from 50% to 25% would reduce steady-state labor hours lost to training by 3%, since new employees need 6 weeks of training.[10] Finally, the capital per employee is comprised of about $5k of desktop IT equipment, $10k of central IT equipment (servers and the network) and $24k of office space (total imputed office rents divided by the total number of employees). Moving employees home for 4 days a week reduces the office space required by 80%, although it increases the desktop IT requirement by 20% (equipment lies idle at home for 1 day a week). On aggregate this reduces capital by 48%. Assuming a coefficient of 1/3 on capital and 2/3 on labor this yields an estimated TFP increase of 30%. Given that the cross-sectional standard-deviation of TFP reported in Foster, Haltiwanger

[10] Training takes 6 weeks, which given a 50% rate of employee turnover, has to be amortized over 2 years, meaning in steady state about 6% of employees are in training. Hence, reducing attrition by 50% reduces training time by 3%.

and Syversson (2008) is 26%, this highlights how differences in the adoption of WFH across firms could potentially account for quite large differences in (measured) productivity.

V.B. Firm learning

The firm learnt four important results from running the formal experiment versus the non-randomized pilot that they had initially been considering. First, they learned that working-from-home improves performance. Without running a formal experiment, their view was that they could have interpreted the drop in treatment performance shown in Figure 3 as a negative treatment effect. The period of the experiment (December 2010 to August 2011) coincided with a business slow-down for CTrip due to a combination of the (predicted) end of Shanghai Expo 2010 and an (unpredicted) increase in competition from other travel agencies. As a result, the *difference* in performance for the treatment group was negative, and is only positive when evaluated as a *difference of differences* against the control group. This highlights the importance of having a well matched (ideally randomized) control group to strip out these kinds of seasonal and competitive effects.

Second, ex ante there was very little discussion of selection effects on employee performance, but by running the experiment and then rolling this out it became clear that allowing employee choice generated a far greater effect than requiring work from home. The impact of working from home is positive, on average, but appears to have a large variance, so that employee choice leads to a much higher effect, as shown in Figure 5.

Third, having the large sample of treatment and control employees allowed the firm to evaluate the impact on different types of employees. Somewhat surprisingly, they found a very homogeneous impact across all types of employees. For example, in Figure 7, we plot the impact on the top half of the treatment versus control distribution and the bottom half of the treatment vs control distribution. To calculate this, both groups were split in half by the pre-experiment median performance and then compared. What we see is a similar improvement in performance for both groups. CTrip's *ex ante* expectation was that the bottom half of employees were the less motivated ones, and they would perform far worse at home. Table A5 shows a similar result that the impact of working-from-home was homogeneous across a range of other characteristics, including gender, commute time, age, prior experience and living arrangements. These results have led the firm to offer working-from-home to all employee groups going forwards rather than any selected sub-samples (such as high-performers), which they were initially intending to target.

Finally, they were surprised by the dramatic drop in attrition that highlighted how many of their employees valued working-from home. They anticipated a reduction, but nothing like the 50% cut they observed.

V.B. Employees' learning

One direct measure of the extent of employee learning is the number of employees who changed their minds about working from home. Figure 2 shows that after the experiment about 50% of the initial treatment and control volunteers changed their minds and decided to work in the office after the end of the experiment, while 10% of the initial non-volunteer group opted to work from home.

We also designed a survey to inquire into employees' evolving views toward the Program from across all 996 airfare and hotel department employees. We administered the same survey with the help of the Ctrip management in November 2010 and August 2011. Employees were asked specifically whether they were interested in participating in the Work-at-Home Program if they were eligible. They could choose from three answers: "yes", "no" or "undecided". We find of the 568 employees that took part in both surveys, that only 303 (53%) maintained their views, while the remaining 47% changed their minds. Of those, 24% went from "yes" or "undecided" to "no", while 12% went from "no" or "undecided" to "yes", with the remainder switching from "yes" or "no" into "undecided".

In follow-up interviews, most of the interviewed employees who had decided they no longer wanted to work from home cited social reasons. Another group who had thought working from home would be attractive found that it was troublesome for the people with whom they lived (often parents), especially if they were called to work outside normal business hours (so that the others were at home). Finally, some who had initially volunteered had ceased to be eligible because of changed living conditions. In reverse, a number of employees saw the success of their peers that worked from home and switched in favor of this.

VI. CONCLUSIONS

The frequency of working from home has been rising rapidly in the US, with over 10% of the work force now reporting regular home working. But there is uncertainty and skepticism over the effectiveness of this, highlighted by phrases like "shirking from home". We report the results of the first randomized experiment on working from home, run in a 13,000 employee NASDAQ-listed Chinese firm, Ctrip Employees who volunteered to work from home were randomized by even/odd birth-date into a treatment group who worked from home four days a week for nine months and a control group who were in the office all five days in the work week. We found a highly significant 13% increase in performance from home-working, of which 9% was from working more minutes of their shift period (fewer breaks and sick days) and 3.5% from higher performance per minute. We found no negative spillovers onto workers who stayed in the office. Home workers also reported substantially higher work satisfaction and psychological attitude scores, and their job attrition rates fell by over 50%.

This experiment highlights how complex the process of learning about new management practices is. For the CTrip, having no precedent in terms of similar Chinese firms that had adopted working from home for their employees led them to run this extensive field experiment. Given their success, other firms are now likely to copy this practice, generating the type of gradual adoption of a new management practices that Griliches (1957) highlighted. More generally, given the large impact of this practice on employee performance – a $2000 per employee reduction in costs and a 30% increase in TFP – this also provides a management practice based explanation for heterogeneous firm performance.

V. BIBLIOGRAPY (to be completed)

Banbury, Simon and Berry, Dianne. (1998). "Disruption of office related tasks by speech and office noise". British Journal of Psychology, 89, 494-517.

Bento, Antonio, Cropper, Maureen, Ahmed Mobarak and "The Impact of Urban Spatial Structure on Travel Demand in the United States" (with A. Bento, M. Cropper, and K. Vinha), *Review of Economics and Statistics*, Vol. 87, No. 3, August, 2005

Bloom, Nick and Van Reenen, John, (2011), "Human resources and management practices", Handbook of Labor Economics, Volume 4, edited by Orley Ashenfelter and David Card.

Bloom, Nick, Tobias Kretschmer and John Van Reenen, (2009), "Work-life Balance, Management Practices and Productivity', in International Differences in the Business Practice and Productivity of Firms, Richard Freeman and Kathy Shaw (eds). Chicago: University of Chicago Press.

Clark and Tellegen (1988)

Council of Economic Advisors (2010), "Work-life balance and the economics of workplace flexibility", http://www.whitehouse.gov/files/documents/100331-cea-economics-workplace-flexibility.pdf

Foster, Lucia, John Haltiwanger and Chad Syverson (2008), "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" American Economic Review, 98(1), 394-425

Gibbons, Robert, and Rebecca Henderson (2012), "What Do Managers Do? Persistent Performance Differences in Seemingly Similar Enterprises,' in Robert Gibbons and John Roberts, eds. *The Handbook of Organizational Economics*, Princeton, NJ: Princeton University Press.

Griliches, Zvi (1957), "Hybrid Corn: An Exploration in the Economics of Technological Change", Econometrica, volume 25 (4), pp. 501-522.

Hamermesh, Daniel (1990), "Shirking or Productive Schmoozing: Wages and the Allocation of Time at Work," Industrial and Labor Relations Review, January

Lazear, Edward, 2000. 'Performance Pay and Productivity', *American Economic Review* 90(5), 1346-1361.

Lazear, Edward, and Paul Oyer (2012), "Personnel Economics," in Robert Gibbons and John Roberts, eds. *Handbook of Organizational Economics*, Princeton University Press, forthcoming.

Lee, David (2008), “Training, wages and sample selection: estimating sharp bounds on treatment effects”. *Review of Economic* Studies, 76(3), pp 1071-1102.

Leibenstein, H. (1966) “Allocative Efficiency vs." X-efficiency," *American Economic Review, 56(3): 392-415*

Maslach, C., & Jackson, S.E. (1981). *Maslach Burnout Inventory*. Research edition. Palo Alto, CA: Consulting Psychologist Press.

Oettinger, Gerald (2012), “The Incidence and Wage Consequences of Home-Based Work in the United States, 1980-2000”, *Journal of Human Resources* forthcoming

Rossi Hansberg, Esteban, Sarte, Pierre-Daniel and Owens, Raymond (2009), “Firm fragmentation and urban patterns”, International Economic Review, Vol 50(1), pp. 143-186.

Syverson, Chad (2011), ‘’What Determines Productivity at the Micro Level?”, forthcoming *Journal of Economic Literature*.

DATA APPENDIX

Appendix A1: Table for different types of workers and their key performance measures

<table>
<tr><td>Types of Workers</td><td>Department</td><td>Key Performance Measures</td><td>Number of Workers</td></tr>
<tr><td rowspan="2">Order Takers</td><td>Airfare</td><td rowspan="2">Phone Calls Answered
Orders Taken</td><td>89</td></tr>
<tr><td>Hotel</td><td>48</td></tr>
<tr><td rowspan="2">Order Placers</td><td>Airfare</td><td>Notifications Sent</td><td>46</td></tr>
<tr><td>Hotel</td><td>Reservation Phone Calls Made</td><td>25</td></tr>
<tr><td>Order Correctors</td><td>Hotel</td><td>Orders Corrected</td><td>36</td></tr>
<tr><td>Night Shift Workers</td><td>Hotel</td><td>Reservation Phone Calls Made
Orders Corrected</td><td>11</td></tr>
</table>

In the analysis, the Order Takers, Order Correctors and Night Shift Workers were grouped together.

Appendix A2: Explanations on the Work Satisfaction Survey

Work Exhaustion: CTrip's in-house psychology counselors used an adapted excerpt from the Maslach Burnout Inventory Survey to measure the emotional exhaustion of the employees from work. The MBI survey was developed by Berkeley psychologist Christina Maslach and Susan Jackson in the 1970s (see Maslach and Jackson, 1981).

Each employee was asked to evaluate his or her "emotional exhaustion" at the end of the work week. The survey contained 6 questions. Each employee was asked to report how often he has felt the way described at work during the week: feel this way every day, almost all the time, most of the time, half of the time, a few times, rarely, never. The survey questions are listed below:

1. I feel emotionally drained from my work.
2. I feel used up at the end of the work day.
3. I dread getting up in the morning and having to face another day on the job.
4. I feel burned out from my work.
5. I feel frustrated by my job.
6. I feel I am working too hard on my job.

Positive and Negative Attitudes: CTrip's in-house psychology counselors used an adapted 16-item Positive and Negative Affect Schedule (PANAS) developed by Clark and Tellegen (1988) to measure the positive and negative attitudes of the employees.

The survey comprised two mood scales, one measuring positive affect and the other measuring negative affect. Each item was rated on a 5-point scale ranging from 1 = *very slightly or not at all* to 5 = *extremely* to indicate the extent to which the employee felt this way the day he took the survey. To evaluate the positive affect, psychologists summed the odd items. In cases with internally missing data (items not answered), the sums were computed after imputation of the missing values: # items on scale / # actually answered, multiplied by the sum obtained from the answered items. A higher score indicates more positive affect, or the extent to which the individual feels enthusiastic, active, and alert. The negative affect is evaluated similarly by summing up the even items.

The 16 items were (1) Cheerful, (2) Jittery, (3) Happy, (4) Ashamed, (5) Excited, (6) Nervous, (7) Enthusiastic, (8) Hostile, (9) Content, (10) Guilty, (11) Relaxed, (12) Angry, (13) Proud, (14) Dejected, (15) Active, (16) Sad.

Appendix A3: Quality did not change in the experiment

	(1)	(2)	(3)	(4)
Dependent Variable	recording grade	recording grade	conversion (z score)	conversion (z score)
Individual FE	No	Yes	No	Yes
Week fixed-effects	Yes	Yes	Yes	Yes
Experiment*Treatment	-0.007	-0.006	-0.026	-0.026
	(0.008)	(0.008)	(0.071)	(0.065)
Treatment	0.000		-0.011	
	(0.005)		(0.091)	
Number of Employees	89	89	135	135
Number of Weeks	87	87	87	87
Observations	5689	5689	9815	9815

Notes: Sample in the first two columns includes 89 order takes in the airfare department (for whom we can obtain recording grade information). The sample in the last two columns includes 135 order takers in airfare and hotels (the group for which conversion rate data exists). Clustered standard errors. *** denotes 1% significance, ** 5% significance and * 10% significance.

Appendix A4. Lack of any obvious cross-sectional Spillover effects

	(1)	(2)	(3)	(4)
Dependent variable	Overall Performance	Overall Performance	Overall Performance	Overall Performance
Sample	Non-experiment	Control	Treatment	Non-experiment + Control
Specification	IV	IV	IV	IV
Treat/Total	-0.221	-0.574	-0.523	-0.263
	(0.398)	(0.392)	(1.039)	(0.357)
Week FE	Yes	Yes	Yes	Yes
No. of Teams	79	59	56	79
Observations	36660	8218	9587	44846
R-squared	0.410	0.359	0.467	0.398

	IV first stage	IV first stage	IV first stage	IV first stage
Sample	Non-experiment	Control	Treatment	Non-experiment + Control
Dependent variable	Treat/Total	Treat/Total	Treat/Total	Treat/Total
Treat/(Treat+Control)	0.253***	0.390***	0.219***	0.264***
	(0.0226)	(0.0295)	(0.0484)	(0.0236)
Week FE	Yes	Yes	Yes	Yes
No. of Teams	79	59	56	79
Observations	36660	8218	9587	44846
R-squared	0.881	0.903	0.891	0.874

Notes: "Treat/total" is the number of employees in treatment divided by the number of employees in each team. A team is composed of 10 to 20 employees who specialize in the same type of tasks and work the same schedule of shifts. Teams typically included treatment and control group members as well as employees not taking part in the experiment. Each team was monitored by the same team leader. "Treat/(Treat+Control)" is the number of employees in treatment divided by the number of employees in treatment and control group within each team. Both "Treat/total" and "Treat/(Treat+Control)" are set to zero before the experiment started on December 6, 2010. "Treat/(Treat+Control)" is fixed at the beginning of the experiment. "Non-experiment", "Control" and "Treatment" refer to employees from each group. The sample includes data from January 1, 2010 to August 15, 2011. Clustered standard errors. *** denotes 1% significance, ** 5% significance and * 10% significance.

Appendix A5. Panel A: Treatment Effects Seem Homogeneous across Characteristics

Performance	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Child	Female	Female w/ Child	Commute >120min	Renter	Young	Short prior experience	Short tenure	Live w/ parents	Live w/ spouse	Live w/ friends	Pre-exper performance
experiment x treat x "characteristic"	0.021	0.033	0.035	0.157	-0.198	-0.151	0.050	-0.085	0.038	-0.020	-0.247	0.024
	(0.169)	(0.123)	(0.195)	(0.142)	(0.140)	(0.127)	(0.127)	(0.125)	(0.134)	(0.166)	(0.245)	(0.100)
experiment x "characteristic"	0.001	-0.061	-0.055	-0.070	0.117	0.025	0.026	0.118	0.009	-0.021	0.266	-0.208***
	(0.130)	(0.087)	(0.175)	(0.090)	(0.107)	(0.091)	(0.092)	(0.089)	(0.101)	(0.113)	(0.207)	(0.077)
experiment x Treatment	0.208***	0.193**	0.204***	0.158**	0.256***	0.296***	0.189**	0.251**	0.186	0.212***	0.226***	0.205***
	(0.066)	(0.080)	(0.065)	(0.079)	(0.074)	(0.100)	(0.093)	(0.099)	(0.114)	(0.065)	(0.065)	(0.060)
Observations	18128	18128	18128	18128	18128	18128	18128	18128	18128	18128	18095	18128
R-squared	0.415	0.415	0.416	0.416	0.416	0.415	0.416	0.415	0.415	0.416	0.419	0.415

Notes: The performance z-scores are constructed by taking the average of normalized performance measures (normalizing each individual measure to a mean of zero and standard deviation of 1 across the sample). The sample includes data from January 1, 2010 to August 15, 2011. "young" equal 1 if an employee is under 24. "Short prior experience" equals 1 if an employee with less than 6 months of experience before joining Ctrip. "Short tenure" equals 1 if an employee has worked in Ctrip for less than 24 month by December 2010. "Pre-exper performance" is the average z-score of performance between Jan 1, 2010 and Oct 1, 2010 for each employee. Clustered standard errors. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 1. Characteristics of employees who volunteer to join WFH

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Sample mean
Children	0.123**		0.075	0.065	0.084	0.090		0.092	0.09
	(0.055)		(0.081)	(0.080)	(0.080)	(0.079)		(0.080)	
Married		0.095**	0.054	-0.002	0.039	0.037		0.040	0.15
		(0.044)	(0.063)	(0.064)	(0.064)	(0.064)		(0.065)	
Cost of commute (Yuan)				0.005**	0.005**	0.005**		0.005**	5.54
				(0.002)	(0.002)	(0.002)		(0.002)	
Bedroom				0.097***	0.089***	0.090***		0.092***	0.60
				(0.032)	(0.032)	(0.033)		(0.034)	
Tertiary education and above					-0.087***	-0.090***		-0.089***	0.41
					(0.032)	(0.032)		(0.032)	
Tenure (months)					-0.002***	-0.003***		-0.003***	24.9
					(0.001)	(0.001)		(0.001)	
Gross wage (Yuan)						-0.003	-0.019	0.032	2872
						(0.001)	(0.017)	(0.023)	
Age								-0.001	23.2
								(0.007)	
Male								0.000	0.32
								(0.035)	
Number of Employees	996	996	996	996	996	996	996	996	996

Notes: The total sample covers all CTrip employees in their Shanghai airfare and hotel departments. Willingness to participate was based on the initial survey in November 2010. Employees were not told the eligibility rules in advance of the survey (i.e.: own room, 6+ months tenure, internet connect etc). Gross wage is calculated as a monthly average of salary from Jan 2010 to Sep 2010 (note that 1 Yuan is about 0.15 Dollars).

Table 2: The performance impact of working from home

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Overall Performance	Phonecalls	Phonecalls	Phonecalls Per Minute	Minutes on the Phone
Dependent Normalization	z-score	z-score	log	log	log
Period: 11 months pre-experiment and 9 months of experiment					
Experiment*Treatment	0.226***	0.263***	0.122***	0.033**	0.089***
	(0.064)	(0.064)	(0.026)	(0.013)	(0.028)
Number of Employees	255	137	137	137	137
Number of Weeks	85	85	85	85	85
Observations	17778	9503	9503	9503	9503

Notes: The regressions are run at the individual by week level, with a full set of individual and week fixed effects. Experiment*treatment is the interaction of the period of the experimentation (December 6th 2010 until August 15th 2011) by an individual having an even birthdate (2nd, 4th, 6th, 8th etc day of the month). The pre period refers to January 1st 2010 until December 5th 2010. The z-scores are constructed by taking the average of normalized performance measures (normalizing each individual measure to a mean of zero and standard deviation of 1 across the sample). Since all employees have z-scores but not all employees have phonecall counts (because for example they do order booking) the z-scores covers a wider group of employees. Minutes on the phone is recorded from the call logs. Standard errors are clustered at the individual level. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 3: Decomposition of the change in labor supply

VARIABLES	(1) Minutes on the Phone	(2) Minutes on the Phone	(3) Minutes on the Phone/ Hours Worked	(4) Hours Worked/ Days Worked	(5) Days Worked
Sample	All	Airfare	Airfare	Airfare	Airfare
Period: 11 months pre-experiment and 9 months of experiment					
Experiment*Treatment	0.089***	0.090**	-0.017	0.068**	0.039**
	(0.028)	(0.044)	(0.033)	(0.028)	(0.015)
Number of Employees	137	89	89	89	89
Number of Weeks	85	85	85	85	85
Observations	9,503	3531	3531	3531	3531

Notes: The regressions are run at the individual by week level, with a full set of individual and week fixed effects. Experiment*treatment is the interaction of the period of the experimentation (December 6th 2010 until August 20th 2011) by an individual having an even birthdate (2nd, 4th, 6th, 8th etc day of the month). The pre period refers to January 1st 2010 until December 5th 2010. Only employees in the Airfare group provides full holiday and leave data so the breakdown by hours and days in the office is only undertaken for this group. Standard errors are clustered at the individual level. *** denotes 1% significance, ** 5% significance and * 10% significance. Minutes on the phone is recorded from the call logs. Hours worked is measured by the phone system log-in and log-out data.

Table 4: The treatment performance also looked good benchmarked against non-experimental and Nantong employees

	(1)	(2)	(3)	(4)
VARIABLES	Overall Performance	Overall Performance	Phone calls	Phone calls
Comparison to Nan Tong				
	Treatment Vs. Nan Tong	Control Vs. Nan Tong	Treatment Vs. Nan Tong	Control Vs. Nan Tong
Experiment*treatment	0.191***		0.241***	
	(0.047)		(0.049)	
Experiment*control		-0.032		-0.032
		(0.048)		(0.044)
Observations	92181	90825	83242	81770
Comparison to Eligible Non-experiment group				
	Treatment Vs. Non-experiment	Control Vs. Non-experiment	Treatment Vs. Non-experiment	Control Vs. Non-experiment
Experiment*treatment	0.209***		0.198***	
	(0.049)		(0.052)	
Experiment*control		-0.021		-0.06
		(0.056)		(0.047)
Observations	48542	47186	31032	30278

Notes: Nan-Tong is CTrip's other large call center, located in Nan-Tong, a city about 1 hour drive outside of Shanghai. This call center also had airfare and hotel departments, and calls were allocated across the Shanghai and Nan Tong call centers randomly. The "Eligible non-experimental group" are the individuals that were eligible for the experiment (own room, 6+ months of tenure and broadband) but did not participate in the two departments in Shanghai. The regressions are run at the individual by week level, with a full set of individual and week fixed effects. Experiment*treatment is the interaction of the period of the experimentation (December 6th 2010 until August 20th 2011) by an individual having an even birthdate (2nd, 4th, 6th, 8th etc day of the month), while Experiment*control is the interaction of the period of the experimentation by an individual having an odd birthdate. All performance measures are z-scores (constructed by taking the average of normalized performance measures, where these are normalizing each individual measure to a mean of zero and standard deviation of 1 across the sample). Standard errors are clustered at the individual level. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 5: Selection Effects

	(1)	(2)	(3)	(4)
Dependent Variable	Log(Phonecalls)	Log(Phonecalls)	Log(Phonecalls)	Log(Phonecalls)
Sample	All	All	All	Balanced
Experiment*WFH	X	0.148***	0.145***	0.143***
	X	(0.041)	(0.037)	(0.049)
Post-Experiment*WFH		0.273***		0.276***
		(0.087)		(0.092)
(0-3 months) Post-Experiment*WFH			0.212***	
			(0.078)	
(3+ months) Post-Experiment*WFH			0.313***	
			(0.090)	
F-test (Exp.*WFH=Post-Exp.*WFH)				
Observations	12653	12653	12653	8294

Notes: WFH here is defined as working-from home at least one day that week. Post-experiment is the period after August 15th 2011 until end of May 2012. Individually clustered standard errors *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 6: Employee switches after the end of the experiment

	(1)	(2)	(3)	(4)
Switch	Home to Office	Home to Office	Home to Office	Home to Office
Performance during the experiment	-0.221		-0.532**	-0.776***
	(0.182)		(0.264)	(0.298)
Performance before the experiment		0.0126	0.442	0.696**
		(0.202)	(0.305)	(0.333)
Married				-0.955*
				(0.499)
Live with parents				-0.629*
				(0.324)
Cost of commute				-0.0340
				(0.0273)
Observations	104	104	104	104

Notes: Sample for returning to the office includes the 104 treatment workers still at CTrip at the end of the experiment in September 2011. Out of the 104 treatment workers, 27 opted to come back to work in the office full-time. Pre-experiment performance is the average of individual weekly performance z-score during the pre-experimental period from January 1st 2010 to December 5th 2010. During experiment performance is the average of individual weekly performance z-score during the post-experimental period from December 6th 2010 to August 15th 2011. The sample for moving home includes the 75 control group employees still in the experiment by September 1st, 2011. Out of 73 control workers, 27 petitioned to work at home, and the company successfully installed the equipment for 25 of them. Robust standard errors. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 7: Employee self-reported work outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Variables:	Satisfaction	General Satisfaction	Life Satisfaction	Exhaustion	Positive Attitude	Negative Attitude
Data source:		Satisfaction survey			Emotion Survey	
Experiment *treatment	0.155***	0.072***	0.168***	-0.564***	0.160***	-0.183***
	(0.052)	(0.021)	(0.047)	(0.168)	(0.040)	(0.058)
Announcement*treatment				-0.102	0.080*	-0.095
				(0.167)	(0.042)	(0.058)
Treatment	-0.015	-0.012	-0.043			
	(0.048)	(0.020)	(0.066)			
Observations	855	855	855	5109	5109	5109

Notes: The satisfaction survey was conducted five times throughout the experimental period: once in early November before the randomization took place and four times after the experiment had started.. See details of survey questions and methodology in Appendix A2. The emotion survey is conducted every week. The first week was conducted in late November 2010, before the experiment begun but after the randomization so that individuals had been informed of their status in the treatment or control groups. All the dependent variables are logged values. The regressions are run at the individual level with a full set of time-dummies. Experiment*treatment is the interaction of the treatment group with the period of the experimentation. Announcement*treatment is the interaction with the treatment group with the period of post-announcement but pre-experiment. Standard errors are clustered at the individual level. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 8. Attrition

	(1)	(2)	(3)	(4)	(5)
Dependent variable	Quit	Quit	Quit	quit	quit
Performance Measure Period	Baseline	Pre-experiment	Post-experiment	Post-experiment	Post-experiment
Sample	Total	Total	Total	Control	Treatment
Performance		-0.315	-1.044***	-1.093***	-0.374
		(0.225)	(0.217)	(0.223)	(0.242)
Performance*Treatment		0.214	0.635*		
		(0.300)	(0.328)		
Treatment	-0.565***	-0.550***	-0.142		
	(0.184)	(0.186)	(0.241)		
Age	-0.114***	-0.107***	-0.0940***	-0.0574	-0.142***
	(0.0332)	(0.0330)	(0.0348)	(0.0469)	(0.0538)
Men	0.190	0.0959	-0.0540	-0.249	0.205
	(0.182)	(0.198)	(0.203)	(0.278)	(0.297)
Married	-0.167	-0.140	-0.290	-0.169	-0.332
	(0.333)	(0.335)	(0.381)	(0.565)	(0.578)
Cost of Commute	0.0288***	0.0291***	0.0296***	0.0305	0.0289**
	(0.0110)	(0.0111)	(0.0111)	(0.0249)	(0.0120)
Children	0.558	0.595	0.930**	0.622	1.259*
	(0.369)	(0.374)	(0.423)	(0.549)	(0.688)
Constant	1.949**	1.795**	1.070	0.298	1.908
	(0.761)	(0.756)	(0.799)	(1.073)	(1.196)
Observations	255	254	254	122	132

Notes: The regressions are all probits at the individual level. The dependent variable is whether the employee quit over the experimental period between December 6th 2010 and August 20th 2011. Pre-experiment performance is the average of individual weekly performance z-score during the pre-experimental period from January 1st 2010 to December 5th 2010. Post-experiment performance is the average of individual weekly performance z-score during the post-experimental period from December 6th 2010 to August 20th, 2011. Performance*treatment is the interaction of the performance measure by an individual having an even birthdate (2nd, 4th, 6th, 8th etc day of the month). Cost of commute is measured at daily level in Chinese Yuan (note that 1 Yuan is about 0.15 Dollars). Standard errors are clustered at the individual level. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 8: Employee survey views before and after the experiment

Rows: Interested in in working from home: August 2011. Columns: Interested in working from home: November 2010

	No	Yes	Undecided	Total
No	71 12.5	59 10.39	79 13.91	209 36.8
Yes	12 2.11	181 31.87	55 9.68	236 41.55
Undecided	17 2.99	43 7.57	51 8.98	123 21.65
Total	100 17.61	295 51.94	173 30.46	568 100

Notes: The total sample covers all CTrip employees in their Shanghai Airfare and Hotel group in November 2010 and August 2011. For the November 2010 survey employees were not told the eligibility rules in advance of the survey (i.e.: own room, 6+ months tenure, internet connect etc). For the November 2011 survey they were told the experiment was being rolled out to the company, but again not what the criteria for this would be.