Expectations-based Reference-Dependence and Labor Supply:

Eliciting Cabdrivers' Expectations in the Field

By Vincent P. Crawford, Miao Jin, Juanjuan Meng, and Lan Yao[1]

Abstract: This paper reports a field experiment on Shanghai cabdrivers' labor supply, analyzing the data using an expectations-based reference-dependent model that allows daily income- and hours-targeting. Our main innovation is to elicit the cabdrivers' income and hours expectations, twice a day. We find that expectations indeed affect labor supply in a way predicted by a reference-dependent model and that income and hours expectations are influenced by their most recent historical average values. Income expectations do adjust within the day, but hours expectations are sticky. Hours targeting has a stronger influence on labor supply than income targeting.

Keywords: expectations-based reference-dependent preferences, loss aversion, income- and hours-targeting, labor supply

***JEL* classification numbers**: D01, D91, J22

[1] Crawford: Department of Economics and All Souls College, University of Oxford, Oxford OX1 4AL, U.K., and Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, U.S.A. (e-mail: vincent.crawford@economics.ox.ac.uk); Jin: National Academy of Development and Strategy, Renmin University of China, Beijing, 100871, China (email: miaojin@ruc.edu.cn); Meng: Guanghua School of Management, Peking University, Beijing, 100871, China (e-mail: jumeng@gsm.pku.edu.cn); Yao: School of Economics, Shanghai University of Finance and Economics, 777 Guoding Rd, Shanghai 200433, China (email: yao.lan@mail.shufe.edu.cn).

1. Introduction

This paper reports a field experiment on Shanghai cabdrivers' labor supply, analyzing the data using an expectations-based reference-dependent model that allows daily income-targeting as in Camerer et al.'s (1997) and Farber's (2005, 2008) econometric analyses, and also income- and hours-targeting as in Crawford and Meng's (2011) econometric analysis and Kőszegi and Rabin's (2006) theoretical analyses. Aside from using new field data to test reference-dependent models, our main innovation is to use China's WeChat messaging app to elicit cabdrivers' expected daily income and working hours, twice a day. Directly eliciting expectations allows us to more accurately identify reference-dependent models of drivers' labor supply and yields valuable new evidence on the dynamics of expectations formation.

Reference-dependent models have played an important role in empirical analyses of workers', consumers', and investors' choice behavior since Camerer et al.'s (1997) analysis of the daily labor supply of New York City cabdrivers.[2] A neoclassical model of labor supply, analogous to a model of consumer demand for income and leisure, predicts a positive elasticity of hours worked with respect to the wage unless there are very large income effects. However, Camerer et al. (1997), taking expected earnings per hour as analogous to the wage, estimate a strongly negative elasticity.[3]

To explain this anomaly Camerer et al. (1997) proposed a model in the spirit of Kahneman and Tversky's (1979) and Tversky and Kahneman's (1991) prospect theory, in which a decision maker has reference-dependent preferences that respond not only to levels of consumption, as in neoclassical

[2] Cabdrivers' labor supply is of particular interest empirically because, unlike most workers in modern economies, many choose their own hours. Another impetus to applications was Kahneman, Knetsch, and Thaler's (1990) experimental analysis of the endowment effect, whereby a person's willingness to accept money for a good he owns exceeds his willingness to pay for it. More recent applications include Odean's (1998), Barberis and Xiong's (2009), and Ben-David and Hirshleifer's (2012) analyses of investors; Oettinger's (1999) study of stadium vendors; Genesove and Mayer's (2001) study of home sellers; Fehr and Goette's (2007) field experiment on bicycle messengers; Post et al.'s (2008) analysis of the game show Deal or No Deal; Pope and Schweitzer's (2011), Lien and Zheng's (2015), and Meng and Weng's (2018) field studies of risky choice; and Li et al.'s (2022) study of insurance sale agents' labor supply response to changes in stock market wealth.

[3] Other studies that find negative wage elasticities include Chou (2002) and Chang and Gross (2014). Studies that find a positive effect of cumulative income on stopping probabilities include Agarwal et al. (2015); Dupas, Robinson, and Saavedra (2020); and Thakral and Tô (2021). By contrast, Farber (2015) analyzes a much larger dataset on New York City cabdrivers and finds nonnegative wage elasticities, and Oettinger (1999), Stafford (2015), Goldberg (2016), Sheldon (2016) and Saia (2017) reach similar conclusions.

consumer theory, but also to changes in consumption measured relative to a reference point, or target. Empirical and experimental studies (summarized in Goette et al. 2020 and Brown et al. 2021) find that most decision makers are loss averse: more sensitive to changes below a target (losses) than changes above it (gains). In Camerer et al.'s (1997) model drivers have daily income targets, which create kinks in their preferences that, with strong enough loss aversion, make realized income tend to bunch around its target. Drivers then tend to work less on days when wages are high, which allows the model to reconcile a negative overall wage elasticity with the normally positive neoclassical incentive effect of a predictably higher wage. Importantly, if drivers are reference-dependent, taking it into account is essential to correctly estimate even the conventional neoclassical determinants of their behavior.

Building on Camerer et al.'s (1997) analysis, Farber (2005, 2008) analyzed a new dataset on New York City cabdrivers, allowing drivers to have daily income targets. In Farber's (2005, 2008) data, as in Camerer et al.'s (1997), earnings per hour is negatively correlated with hours. A challenge in testing reference-dependent models in the field is that targets are normally unobservable. Treating targets econometrically as latent variables, Farber finds that a reference-dependent model fits better than a neoclassical model. However, his estimates of the targets are unstable, which he argues precludes a useful reference-dependent model.

In a theory paper inspired by Camerer et al.'s (1997) and Farber's (2005, 2008) analyses, Kőszegi and Rabin (2006) adapt Kahneman and Tversky's (1979) model of reference-dependent preferences to economic applications. Kőszegi and Rabin (2006) allow preferences to reflect reference-dependent "gain-loss" utility as well as neoclassical consumption utility, with a separate target for each good. Unlike Kahneman and Tversky (1979), who take no definite position on how targets are determined, Kőszegi and Rabin (2006) close their model by setting an agent's target equal to the rational expectation of his subsequently chosen consumption, good by good. Such an expectations-based reference-dependent model can reconcile negative overall wage elasticities with the neoclassical intuition that predictably higher wages normally increase labor supply: For anticipated changes in income or hours, gain-loss utility drops out of the model, leaving only consumption utility, which reproduces the neoclassical intuition. But for unanticipated changes, loss aversion makes drivers' choices bunch around the targets, which can yield a negative elasticity.

Crawford and Meng (2011) adapt Kőszegi and Rabin's (2006) model to reconsider Farber's (2005, 2008) analyses econometrically, using Farber's data. As Kőszegi and Rabin's (2006) model suggests, they allow daily hours targets as well as Farber's (2005, 2008) daily income targets. Crawford and Meng (2011) address the unobservability of targets by implementing Kőszegi and Rabin's (2006) rational-expectations view, assuming that in steady state, the expectations that determine drivers' targets equal their natural sample analogs, driver by driver. This allows their econometric analysis to avoid the instability of Farber's (2008) estimated latent targets and most of the other difficulties that led him to doubt the usefulness of a reference-dependent model. As Kőszegi and Rabin (2006) suggested, Crawford and Meng's (2011) estimated model reconciles Farber's negative wage elasticity with the neoclassical intuition that predictably higher wages normally increase labor supply.

Kőszegi and Rabin's (2006) analysis inspired other experimental and observational studies, which address the unobservability of expectations in several ways. Abeler et al. (2011), Ericson and Fuster (2011), Banerji and Gupta (2014), Sprenger (2015), Karle, Kirchsteiger, and Peitz (2015), Song (2016), and Goette et al. (2020) report laboratory experiments that vary subjects' implied rational expectations regarding probabilities, finding that subjects' choices change in ways that generally support expectations-based reference-dependence. However, the experimental support is not unequivocal (Heffetz and List, 2014; Gneezy et al., 2017); and some of it is indirect in that it requires assumptions about how reference points are formed, or rests on qualitative inferences. Observational studies either treat drivers' targets as latent variables (Camerer et al., 1997; Farber, 2005), use sample proxies for rational-expectations targets (Crawford and Meng, 2011), or structurally estimate the targets (Farber, 2008; Thakral and Tô, 2021).[4] The

[4] In a nonparametric analysis of reference-dependent models of consumer demand, Blow and Crawford (2023) show that either observing or precisely modeling reference points is essential for the model to have *any* nonparametrically refutable implications. Thus, treating targets as latent variables largely reduces tests of a reference-dependent model to tests of functional form assumptions.

debate remains on the extent to which the evidence supports expectations-based reference-dependent models.[5]

This paper adds new evidence to this debate from a field experiment on Shanghai cabdrivers' labor supply. Our data analysis, like Crawford and Meng's (2011), allows the possibility that drivers have expectations-based reference-dependent preferences with daily income- and hours-targeting. Our main innovation is that, instead of treating targets as sample proxies for rational expectations or latent variables, we elicit drivers' expectations on daily income and hours directly, twice a day, using China's messaging app WeChat (https://www.wechat.com/en/).[6] This allows us to dispense with Crawford and Meng's rational expectations assumption and yields a more accurate model of drivers' decisions. In addition to providing new evidence on expectations-based reference-dependent models of labor supply, our experiment yields new evidence on the dynamics of drivers' expectations formation. This complements the large literature on the dynamics of expectations in other contexts.

Our field experiment can be summarized as follows. We carried out two within-subject interventions on cabdrivers from May to September, 2018: Expectation Treatment and Neutral Treatment. In the Expectation Treatment we elicited drivers' expectations about income and working hours of the current shift twice a day with payments. In the Neutral Treatment, to control for possible effects of the elicitation on drivers' working patterns, we asked neutral questions about drivers' background information in a similar format. Neither kind of elicitation significantly changes drivers' shift-level wage elasticities or their trip-level relationships between cumulative income per hour and stopping decisions. This suggests that drivers' working patterns were comparable to what they would have been without our interventions.

We next estimate whether drivers' elicited expectations of income and working hours serve as reference points in determining their trip-level stopping decisions. Both morning and afternoon expected hours significantly affect stopping decisions. Afternoon expected income also has a significant effect, but the

[5] Brandon et al. (2023) and Dupas, Robinson, and Saavedra (2020) discovered that randomized windfall income had no impact on labor supply in their field experiments. This evidence contradicts the predictions of the reference-dependent model if one assumes that drivers' mental accounting equates windfall income with trip income.
[6] We use the terms "expectation" and "target" interchangeably in this paper.

effect of morning expected income is statistically insignificant. Splitting the sample by whether the realized wage (income per hour) between the two daily elicitations is higher or lower than the morning expected wage also suggests that elicited expectations affect stopping decisions in the asymmetric way that is characteristic of expectations-based two-target models like Crawford and Meng's (2011): That is, when the realized wage is lower (respectively, higher) than expected, the stopping probability is significantly driven by income (respectively, hours) expectations.

We then investigate the formation and dynamic adjustment of expectations. We find that income and hours expectations are significantly affected by their most recent history, that is, the average income and hours in the past five shifts with the same day of the week. Individual-level factors such as daily consumption and daily cab rental fee do not significantly affect expectations formation. With regard to dynamic adjustment, when the wage realized between the two daily elicitations is below (respectively, above) the morning expected wage, drivers do update their income expectations downward (upward). However, in the analogous situations for hours, there is no significant adjustment in hours expectations. The fact that hours expectations are more stable than income expectations, combined with their stronger influence on drivers' stopping decisions, suggests that hours targeting is a more important factor than previously thought and that a two-target model such as Crawford and Meng's (2011) improves on the income-targeting models used in almost all previous work. Notice that there is a distinction between the impact of hours targeting and the direct effect of hours from a neoclassical perspective: The former is identified through the driver's reactions to unexpected changes in hours relative to the hours target, while the latter is identified through expected changes. Behaviorally, hours targeting results in zero wage elasticity, while the neoclassical effect is anticipated to yield positive wage elasticity in our setting.[7]

[7] The estimated effect of hours targets in our study is unlikely to be influenced by scheduled obligations or plans. This is primarily because the majority of stopping hours in our sample occur after 10pm, a time when obligatory activities are less likely.

Finally, comparing models with directly elicited expectations versus expectations proxied via historical sample averages (as in Crawford and Meng 2011) suggests that elicited expectations better explain drivers' labor supply behavior, so expectations elicitation adds value.

The rest of the paper is organized as follows. Section 2 describes our experimental design. Section 3 summarizes the data and reports preliminary statistical tests. Section 4 presents econometric estimates, focusing on how drivers' elicited expectations of income and hours affect their stopping decisions. Section 5 reports econometric estimates on how drivers form and update their expectations. Section 6 concludes.

2. Experimental Design

Our experiment studies the labor supply decisions of the cabdrivers in a taxi leasing company in Shanghai in 2018. The drivers lease their cabs annually from the company. Within the year they choose their working hours freely and collect the fare income, paying a fixed monthly rental fee and paying for fuel. Our experiment began with a pre-experiment survey and then elicited expectations. The time span was April 24 to September 30, 2018, with expectations elicitation from May 5 to September 30.[8]

2.1 Pre-Experiment Survey

The pre-experiment survey was conducted at drivers' regular monthly meetings. We surveyed their background information, such as monthly consumption, monthly rental fee for the cab, and normal working days of the week; and preference measures such as degrees of loss aversion and time preference, and narrow bracketing.[9] Appendix B lists the survey questions translated into English. We also obtained drivers' contact information via their WeChat accounts.[10] For completing the survey, we paid each driver a fixed fee of 30 CNY, or US$4.27.

---

[8] Because not every driver showed up for their monthly meetings, we conducted the survey with three waves of drivers, which finished the survey on April 24, May 22 to May 25, and June 20 to June 24, respectively. The five-month elicitation period was due to our desire not to interrupt drivers too frequently with our questions.

[9] We surveyed drivers' preference measures to investigate how they interacted with expectations formation. We found no significant effects, and therefore do not discuss those measures in the paper.

[10] WeChat is a messaging app in China which provides services, including free messaging, free calling, transferring money between users, etc. (https://www.wechat.com/en/). The messaging and calling services are like those offered

2.2 Eliciting Expectations

Expectations elicitation was done in the field via China's messaging app WeChat. To more fully assess how elicitation affects drivers' labor supply choices, we adopt a within-subject design in which each driver was given two treatments but on different, randomly selected days, an Expectations Treatment or a Neutral Treatment.

In an Expectations Treatment, we elicited a driver's expectations about the current shift's income and working hours, twice a day. Specifically, at 4:00 am ("morning expectations") we asked "How much do you expect to earn today? What time do you expect to stop work today?"; and at 6:00 pm ("afternoon expectations") we asked the same questions, prefaced by "According to your work so far,…".[11] In the Neutral Treatment, we asked questions unrelated to expectations or labor supply, such as "What are your eating habits?". In each case, a driver is paid 10 CNY, about US$1.40, for each answered question, immediately via WeChat transfer. Comparing drivers' choices across the two treatments reveals whether eliciting drivers' expectations has an effect on their choices beyond that of any intervention that makes them feel observed and obtain a small windfall income.

Each driver experienced the Expectations Treatment on approximately 10 days and the Neutral Treatment on approximately 5 days, with treatments separated by 6 days on average. Thus, each driver experienced days with one or the other treatment and days with neither. Each driver's Expectations Treatments covered all seven days of the week to assess day-of-the-week effects, and drivers' Neutral Treatments were randomly allocated across all days of the week. In the whole sample, the two treatments were evenly distributed over days of the week (Figure A1 in Appendix A).

3. Data and Preliminary Statistical Tests

by Facebook's and WhatsApp and the money transfer service is similar to PayPal. The WeChat app covers more than one billion users and is in common use in China.

[11] We chose those times because in our sample drivers' shifts, on average, start at 7:10 am and end at 11:40 pm. The average start/end time in the driving records data are very similar to those in the survey data; the drivers on average state that they start at 6:10 am and end at 11:50 pm.

3.1. Data

We elicited expectations for 147 of the company's drivers during the period May 5 to September 30, 2018. For those drivers, we also obtained trip-level data, which are collected automatically via electronic devices in the cabs, from the company. These data include driver and car identification numbers and trip-by-trip fares, start times, end times, kilometers traveled with passengers in the cab, and kilometers traveled between the last trip and the current trip (searching for passengers). Tipping drivers is not customary in China. Unfortunately, our data do not include location information. The electronic device computes each trip's fare based on a standard city rate for the sample period in 2018: 14.00 CNY (US$2.00) for the first 3 kilometers, plus 2.50 CNY (US$0.36) for each four-minute block of waiting time, plus 2.50 CNY (US$0.36) for each additional kilometer up to 15 kilometers and 3.75 CNY (US$0.53) for each additional kilometer above 15 kilometers. Between 11 pm and 5 am, there is a surcharge of 30%. In addition, we collected data on the weather: temperature and precipitation data for each 30-minute period from the National Climatic Data Center (NCDC) and China Meteorological Administration respectively.

We define a shift as a sequence of consecutive trips that are not more than six hours apart.[12] To remove outliers, we exclude shifts with more than 24 working hours; and we remove drivers for whom more than 20% of their shifts exceeds 24 hours. These exclusions leave us with what we will call a "working sample" of 131 drivers, who took 241,761 trips in 10,137 shifts.

[Table 1 here]

Table 1 gives summary statistics for our working sample. Panel A gives shift-level summary statistics. On average drivers work 16.5 hours per shift and earn 972.2 CNY (US$138.90). Their average wage (total income divided by working hours) is 61.0 CNY (US$8.70) per hour. Within a day, across all drivers, the hourly wage is significantly positively autocorrelated, with an autocorrelation coefficient oof 0.075 (p =

[12] If there is more than a six-hour gap between two trips, we treat the first trip as the end of one shift and the second trip as the beginning of the next. Farber (2005, 2008) and Crawford and Meng (2011) used data based on manual records of the trips that defined "shift" in its original form. Farber (2015) and Thakral and Tô (2021) used data collected electronically, making it necessary to define a shift. This paper follows their definitions.

0.000).[13] Drivers' average morning income expectation is 969.0 CNY (US$138.40) and their average afternoon expectation is 962.3 CNY (US$137.50). Drivers' average morning and afternoon hours expectations are both 17.7 hours.

Panel B gives driver-level summary statistics. On average, a driver's monthly consumption is 5422.1 CNY (US$774.60) and his monthly rental fee is 5736.8 CNY (US$819.50). Panel C gives trip-level summary statistics, mainly weather conditions.

Our data analysis proceeds as follows. In the rest of this section, we test whether our experimental intervention affects drivers' behavior. Ideally, there should be no effect or a small effect. In Section 4 we estimate how drivers' elicited daily income and hours expectations affect their stopping decisions, following Farber's (2005, 2015) and Crawford and Meng's (2011) econometric strategy. In Section 5 we estimate how drivers formed and updated their expectations.

3.2 Preliminary Statistical Tests

In this section, we perform two sets of preliminary statistical tests. First, we test whether our experimental intervention affects drivers' working patterns. One concern is whether expectation elicitation by itself systematically changes their choices, possibly from neoclassical to reference-dependent, simply by reminding them of the possibility of having income and hours targets. We then test whether drivers' elicited expectations contain meaningful information about their expectations.

3.2.1 Intervention Effects on Shift- and Trip-level Working Patterns

We test for the intervention effects by comparing drivers' shift- and trip-level outcomes with and without interventions. Recall that in our design, a given driver experiences days with the Expectations Treatment, days with the Neutral Treatment, and days with no intervention. We therefore divide the sample, by driver-shifts, into those three categories and check for differences in drivers' working patterns.

First, we estimate the effect of the intervention on wage elasticity using the following specification:

[13] The autocorrelation coefficient is consistent with the findings in Farber (2005), which is 0.0687** (significant at the 5% level). We also calculate the within-day autocorrelation coefficient of the market hourly wage (the median hourly wage across all drivers). The autocorrelation coefficient of 0.765 ($p = 0.000$) is larger than that in Camerer et al. (1997), which is 0.493 (standard error 0.092).

$$lnH_{is} = \beta_0 + \beta_1 lnW_{is} + \beta_2 lnW_{is}NT_{is} + \beta_3 lnW_{is}ET_{is} + \beta_4 NT_{is} + \beta_5 ET_{is} + X_{is}\gamma + \mu_i + \epsilon_{is} \quad (1)$$

where lnH_{is} is driver i's log working hours in shift s, and lnW_{is} is his log wage in this shift. NT_{is} and ET_{is} are indicators for Neutral Treatment and Expectation Treatment, respectively. Following the literature (Camerer et al. 1997; Farber 2005, 2015), the control variables X_{ist} include day-of-the-week fixed effects, day-of-the-month fixed effects, month fixed effects, shift ending time, holiday, weather, and other variables. μ_i denotes driver fixed effects. Our analyses below share the same set of controls, except as indicated. Standard errors are clustered at the individual driver level.

Column (1) of Table 2 reports results from the OLS regression. The estimated coefficient of log wage is significantly negative, which is in line with the literature (Camerer et al., 1997; Farber, 2005, 2015). More importantly, the estimated coefficients of the interaction terms between treatment indicators and log wage are statistically insignificant, suggesting that answering expectations or neutral questions does not alter drivers' wage elasticity of labor supply. In column (2), we use instrumental variables (IV) to alleviate bias due to measurement error. Following Farber (2015), we use the average log wage of other drivers on the same day as an instrument for own log wage. The IV estimate of the labor supply elasticity is positive but insignificant and the interaction terms of the treatment indicators and log wage are again insignificant. Importantly, both OLS and IV estimates suggest that our intervention does not significantly alter drivers' wage elasticity.

Next, following Farber (2005, 2015) and Crawford and Meng (2011), we examine how drivers' stopping decisions respond to cumulative income and hours at the trip level, and whether such effects are altered by our intervention, using the following specification:

$$d_{ist} = \beta_0 + \beta_1 h_{ist} + \beta_2 y_{ist} + \beta_3 h_{ist}NT_{is} + \beta_4 h_{ist}ET_{is} + \beta_5 y_{ist}NT_{is} + \beta_6 y_{ist}ET_{is} + \beta_7 NT_{is} + \beta_8 ET_{is} + X_{ist}\gamma + \mu_i + \epsilon_{ist} \quad (2)$$

where d_{ist} is driver i's stopping decision at the end of trip t in shift s, h_{ist} is cumulative hours, and y_{ist} is his cumulative income. NT_{is} and ET_{is} are indicators for Neutral Treatment and Expectation Treatment, respectively. X_{ist} denotes control variables and μ_i denotes driver fixed effects. Column (3) of Table 2 reports the results. In the absence of intervention, we find an overall insignificantly positive effect of cumulative income and a significantly positive effect of hours on the probability of stopping, which is consistent with most findings in the literature (Farber 2005, 2015; Crawford and Meng 2011). Importantly, the interaction terms between treatment indicators and cumulative income (or hours) are statistically insignificant, again suggesting that our interventions do not alter trip-level working patterns.

[Table 2 here]

In addition, following Thakral and Tô (2021), we employ local linear regression to estimate the time-varying effects of cumulative income on drivers' stopping decisions and the effects of our intervention:

$$d_{ist} = f(h_{ist}) + \sum_{l} \beta^l(h_{ist}) y_{ist}^l + \sum_{l} \rho^l(h_{ist}) y_{ist}^l NT_{is} + \sum_{l} \theta^l(h_{ist}) y_{ist}^l ET_{is} + \delta(h_{ist}) NT_{is}$$

$$+\sigma(h_{ist}) ET_{is} + X_{ist}\gamma(h_{ist}) + \mu_i(h_{ist}) + \epsilon_{ist} \quad (3)$$

where d_{ist} is driver i's stopping decision at the end of trip t in shift s. h_{ist} is cumulative hours. As in Thakral and Tô (2021), we divide cumulative income into income earned in different hours of the shift. y_{ist}^l denotes income earned in hour l. To examine the effects of our interventions, we then interact income earned in each hour with intervention indicators NT_{is} and ET_{is}, respectively. X_{ist} denotes control variables and μ_i denotes driver fixed effects. We test the coefficients of the interaction terms between an intervention indicator and the income earned across all hours jointly. As shown in Appendix Table A2, the p-values from the joint tests indicate no statistically significant differences between the Untreated, Neutral Treatment, and Expectation Treatment groups over the course of the shift. Therefore we do not observe that our treatment affects labor supply pattern even from the most recent non-parametric point of view in Thakral and Tô (2021).

3.2.2 Intervention Effect on Shift-level Income and Working Hours

We next examine whether our intervention affects the level of shift income or working hours.

$$Y_{is} = \beta_0 + \beta_4 NT_{is} + \beta_5 ET_{is} + X_{is}\gamma + \mu_i + \epsilon_{is} \quad (4)$$

where Y_{is} is driver i's total income or working hours in shift s. Table 3 reports the regression results. The estimates show that the Expectations and Neutral Treatments increase shift income by 24.1 CNY (US$3.40) and 17.6 CNY (US$2.50) per shift, respectively. These effects are small (2.48% and 1.81% of the sample mean of 972.2 CNY) but statistically significant. We find similarly significant effects on working hours. The Expectations and Neutral Treatments increase hours by 0.5 and 0.4 respectively (3.03% and 2.42% of the sample mean of 16.5 hours). F tests suggest that these effects do not differ significantly between the Expectations and Neutral Treatments. These effects appear to come simply from being asked questions or the feeling of being observed, not from the specific content of the expectations elicitation.

[Table 3 here]

Overall, considering the shift-level and trip-level analyses, our intervention does not significantly change the wage elasticity or the relationship between cumulative income (or hours) and stopping probability. Eliciting expectations does not convert drivers from neoclassical to reference-dependent, or vice versa.

Although we pay drivers for answering the expectations questions, those questions cannot be properly incentivized, so it is important to check whether the elicited expectations are credible. We report some observations about the expectations. First, the expectations vary across days of the week in a pattern similar to that of the realized income and hours (Figure 1). For example, both expected and realized income, as well as working hours, reach a peak on Friday and decline over the weekend. This suggests that drivers' elicited expectations are at least qualitatively consistent with rationally anticipating variations across days. Second, the morning (respectively, afternoon) expected income is positively correlated with realized income, with a correlation coefficient of 0.474 (0.548). And morning (afternoon) expected hours is positively correlated with realized hours, with a coefficient of 0.769 (0.773). Overall, our elicited expectations data appear to contain meaningful information about drivers' expectations.

4. Elicited Expectations and Drivers' Stopping Decisions

This section examines whether and how our elicited income and hours expectations affect drivers' stopping decisions. We stress that the effect of hours targeting is distinct from the neoclassical direct effect of hours on preferences. The effect of hours targeting is identified by the driver's response to unanticipated changes in hours, while the neoclassical effect regards anticipated hours. In terms of behavioral patterns, the former leads to zero wage elasticity while the latter is expected to generate a positive wage elasticity.

We restrict our working sample to trips in shifts that had the Expectations Treatment. Our econometric strategy follows Farber (2005, 2015) and Crawford and Meng's (2011, Section II.A)): [14]

$$d_{ist} = \beta_0 + \beta_1 h_{ist} + \beta_2 1_{\{h_{ist} > H_{is}\}} + \beta_3 y_{ist} + \beta_4 1_{\{y_{ist} > Y_{is}\}} + X_{ist}\gamma + \mu_i + \epsilon_{ist} \quad (5)$$

where d_{ist} is driver i's stopping decision at the end of trip t in shift s. h_{ist} is cumulative hours, H_{is} is driver and shift specific hours target, y_{ist} is his cumulative income, and Y_{is} is driver and shift specific income target. X_{ist} denotes control variables, and μ_i denotes driver fixed effects.

According to Crawford and Meng's (2011) multi-target model, not reaching the income target or exceeding the hours target are both coded as "losses", hence loss aversion creates a tendency for stopping probabilities to increase once each target is reached. We find that reaching either the morning or afternoon hours target significantly increases stopping probabilities. Reaching the afternoon income target significantly increases stopping probabilities, but the effect of reaching the morning income target is insignificant.

To provide a stronger test of the role of expectations-based reference points, we also do a split-sample estimation following Crawford and Meng (2011), according to whether the income or hours target was reached first. The two-target model suggests that the income target will play the major role in one subsample and the hours target in the other one, i.e., the effects of income and hours targets will be asymmetric. The direction of the exact asymmetry depends on whether drivers tend to stop at the first or second target reached on a given day, as well as whether the income or hours target is reached first. Because this asymmetric pattern is unique to the two-target model, estimation with split samples is a stronger test of the role of

[14] Brandon et al. (2023) point out two concerns with specifications like Farber's (2005, 2015) and Crawford and Meng's (2011). First, drivers' labor supply decisions may be influenced by bonus income (tips and other incentives) as well as meter income, but in Farber's and Crawford and Meng's studies tips are unobserved or partly observed. Second, the control variables should distinguish between hours and kilometers with and without passengers. Our study avoids most of those problems, because tipping is not customary in China and we control for the ratio of kilometers traveled with passengers to total kilometers; but our expectations data do not distinguish between hours with and without passengers.

expectations as the reference point. Overall, the subsample results are consistent with predictions of a two-target model in which drivers stop at the second target they reach, which is compelling evidence of expectations-based reference dependence.

We present the whole sample and split-sample analysis in detail below.

4.1 Whole Sample Analysis

In the whole-sample analysis, we add the morning and afternoon sets of expectations separately (columns (1) and (2)) and together (column (3)) in the estimation reported in Table 4. When either morning or afternoon targets are added separately, there is a significant jump in the probability of stopping when the corresponding target is reached, and the magnitude of the jump is larger in the case of the hours target. Specifically, there is a 4.8 (5.1) percentage point increase in the odds of stopping when the morning (afternoon) income target is reached, and a 21.6 (21.3) percentage point increase after the morning (afternoon) hours target is reached. The difference in the increase between reaching the morning (afternoon) income target and reaching the hours target is statistically significant. When morning and afternoon targets are included simultaneously, estimates in column (3) suggest that both the morning and afternoon hours targets have significant effects, with a cabdriver being 14.3 (9.3) percentage points more likely to stop work when the morning (afternoon) hours target is reached. However, only the afternoon income target exerts a significant effect, leading to a 3.7 percentage point increase in the probability of stopping when it is exceeded, but the morning income target does not show a significant effect anymore. The results show that cabdrivers care more about the afternoon than morning income expectations and consider both the morning and afternoon hours expectations in their stopping decisions. This aligns with the findings in Section 5 that income expectations adjust over time while hours expectations do not, explaining why afternoon, rather than morning, income expectations significantly affect stopping decisions.

Aside from the effects of reaching targets, the estimated coefficients of cumulative income are insignificant, but those of cumulative hours are significantly positive across specifications. The overall results are very similar to those of Crawford and Meng (2011), indicating that the effect of cumulative income mainly comes from whether it exceeds the target rather than from its level, but cumulative hours have both a neoclassical level effect and a separate, reference-dependent effect of reaching the target.

[**Table 4 here**]

Finding 1. Both morning and afternoon hours targets have significant effects on the stopping decisions. Although the afternoon income target has a significant effect, the effect of the morning income target is not significant.

4.2 Split-Sample Analysis

Finding 1 provides some general evidence on the significant effect of the elicited expectations. However, to understand more deeply whether expectations affect stopping decisions in the same way as reference points, we resort to an analysis that splits the sample according to whether the income or hours target is likely to be reached first. As mentioned at the beginning of Section 4, the two-target model in Crawford and Meng (2011) predicts a unique asymmetric pattern in the resulting subsamples: the income target will significantly affect the stopping decision in one subsample and the hours target in the other one. For instance, if drivers tend to stop at the second target reached, as seen in Crawford and Meng (2011), we expect to see a significant effect of reaching the income (or hours) target on the stopping decision in the subsample where the income (or hours) target is likely to be reached second. The asymmetric pattern still exists if drivers tend to stop at the first-reached target, but with the direction of the asymmetry reversed. The asymmetry across subsamples is strong evidence for the role of expectations as the reference point.[15]

[15] If preferences were homogeneous, as Farber (2005)'s, Crawford and Meng (2011)'s, and our model assume, drivers' stopping probabilities would either all tend to be more strongly influenced by the first target reached on a given day, or all by the second. In our estimates drivers' stopping probabilities happen to be more strongly influenced by the second target a driver reaches on a given day than by the first. Because on good days high early income makes a driver reach his income target before his hours target, working hours has a strong and significant positive effect on the stopping probability. On bad days when early income is low this pattern is reversed, and it is income that has a strong and significant positive effect on the stopping probability. (If, by contrast, drivers' stopping probabilities were more strongly influenced by the first target a driver reaches on a given day than by the second, this entire pattern would be reversed.) Thus the pattern of significance in our results is one of the two possible patterns that are characteristic of a two-target reference-dependent model with homogeneous preferences, and as such is powerful evidence for our reference-dependence model. (With heterogeneous reference-dependent preferences other patterns of significance are logically possible, but more "contrived" and so less plausible. In a neoclassical model it is approximately irrelevant whether early income is unexpectedly high or low.)

The specific way of splitting is as follows: we define the morning expected wage as the morning income expectation divided by the hours expectation, and define the early realized wage as the realized income divided by hours between the period of the morning and afternoon expectation elicitations. Then we look at only trips after the afternoon expectation elicitations, divide them into subsamples according to whether the early realized wage is higher or lower than the morning expected wage and do analysis in each subsample. We can also compare the early realized wage with the afternoon expected wage to split the sample, and the results are robust to that alternative.[16]

Before considering the regression, let us first look at the pattern graphically. We first graph the levels of the expected and actual income and hours by day of the week (Figure 1) and its subsample patterns (Figure 2). Figure 1 shows a clear co-movement between expected and realized income, and the same is true for hours, which we discussed in Section 3.2. Besides this, we also find that the level of income expectation is generally lower than that of realization, while the expectation and realization of working hours have similar magnitudes.

[**Figure 1 here**]

Figure 2 graphs the subsample patterns. If early realized wage is lower than the morning expected wage (Panel A of Figure 2), the hours target is likely to be reached before the income target. Regardless of whether drivers stop at the hours target that is reached first or the income target that is reached second, actual working hours will be no lower than expected on average; and actual income no higher than expected on average. The sub-sample pattern is consistent with this prediction. In Panel A, the average realized working hours exceeds its expectation, but average realized income is lower than expected, regardless of whether the morning or afternoon expectations are used in comparison. In the other case, if early realized wage is higher than the morning expected wage (Panel B of Figure 2), the income target is likely to be

[16] To avoid endogeneity problems, we need to separate the data used to split the sample from that used in the regression. That is why we only focus on trips after the afternoon expectation elicitations. Because both the morning and afternoon expectations can potentially affect the final stopping decisions, we can split the sample by comparing the early realized wage to either the morning or afternoon expectations. However, it is possible that the afternoon expectations will absorb any wage "shock" early on the day so that the sample splitting is less effective.

reached before the hours target, and the prediction is reversed. Indeed, the patterns in Panel B show that the average realized working hours are very close to expectations and that realized income is larger than expectation. Overall, the differences between realizations and expectations in the two subsamples are highly consistent with the predictions of two-target model.

[**Figure 2 here**]

To provide formal statistical evidence, we further analyze the effects of income and hours targets (both morning and afternoon) on cabdrivers' stopping decisions in the two subsamples in columns (4) and (5) of Table 4. Reaching the morning hours target always has a significant positive effect on the stopping probability, regardless of whether realized wage is higher or lower than the morning expectations. However, in column (4) when the early realized wage is lower than the morning expected wage (so that the income target is likely to be reached second), the afternoon income target has a significant impact on the stopping decision, but the effect of the afternoon hours target is insignificant. By contrast, in Column (5) when the early realized wage is higher than the morning expected wage (so that the hours target is likely to be reached second), it is the afternoon hours target rather than the afternoon income target that significantly affects the stopping decision. Overall there is indeed an asymmetric pattern in terms of whether the income or hours targets affect the stopping decision in each subsample, which is consistent with a two-target model. Further, the exact direction of the asymmetry suggests that drivers in our sample also tend to stop at the second-reached target, which is not a necessary prediction of the model but consistent with the sample of New York City cabdrivers in Crawford and Meng (2011).

Finding 2. When the early realized wage is lower (respectively, higher) than the morning expected wage, it is the afternoon income (hours) target rather than the afternoon hours (income) target that has a significant effect on the stopping decisions. This finding provides evidence for the role of expectations as reference point based on a two-target model in which drivers tend to stop when they reach the second target.

Put together, the estimates in the whole sample and subsample analysis suggest that our elicited income and hours expectations play a significant role in determining drivers' stopping probabilities, as suggested by Kőszegi and Rabin's (2006) and Crawford and Meng's (2011) two-target model. Instead of following

Crawford and Meng in using sample averages to proxy the expectations that in these models determine drivers' reference points, we confirm that their elicited expectations serve as reference points in determining labor supply decisions. In the next section, we analyze drivers' expectations formation and the dynamics of their updating in detail.

5. Expectations Formation and Adjustment

In this section, we first analyze what factors drive the formation of drivers' morning expectations, and then study how their expectations adjust between the two daily elicitations.

5.1 Morning Expectations

Crawford and Meng (2011) and Agarwal et al. (2015) assume that income and hours targets are determined by rational expectations, proxied by day-of-the-week historical averages. With the elicited expectations data, we can test this assumption directly. To do this, we construct the variable *Historyincome1to5* (*Historyhour1to5*) to be the average shift income (working hours) of the past five shifts for the same driver and the same day of the week. We also construct similar variables for the past six to ten shifts. In addition to the historical average, some studies also mention that consumption and cost of operation also affect income reference point (Camerer et al., 1997; Dupas, Robinson, and Saavedra, 2020), so we include these variables from the pre-experiment survey.

We estimate the effects of the above factors on morning elicited expectations of income and hours and report the results in Table 5. Separately, and within each category, we either control for individual-level fixed effects or the individual-level characteristics mentioned above, such as daily consumption and daily cab rental fees. Regardless of which controls are used, the most recent history, i.e., the average income in the past five shifts on the same days of the week of the same driver, significantly and positively affects expectations of current income; and the conclusion for hour is also the same.

[**Table 5 here**]

Finding 3. Income and hours expectations are significantly affected by cabdrivers' most recent experience of realized income and hours. Other individual-level factors, including daily consumption and daily cab rental, do not show significant effects on expectations formation.

The positive associations between expectations and historical outcomes generally support the assumptions made in Crawford and Meng (2011) and Agarwal et al. (2015) in analyzing expectations-based reference points in labor supply decision. These results are also consistent with findings that expectations are affected by historical outcomes and experiences in other settings (Malmendier and Nagel, 2011, 2016; Greenwood and Shleifer, 2014; Kuchler and Zafar, 2019).

5.2 Expectation Adjustment

We next study the expectations adjustment based on two elicitations of the expectations within shift. To show the general pattern, Figure 3 presents the scatterplots of the morning expectation against the afternoon ones, for income and hours separately. The figures show that the difference between the fitted slopes and the 45-degree line is small in magnitude but statistically significant, indicating small but significant adjustments in expectations. In fact, the data on elicited expectations show that, in 37.68% (26.16%) of the shifts, drivers adjust their income (hours) expectations.

[**Figure 3 here**]

To understand the detailed adjustment pattern, we pool the morning and afternoon expectations data, put expectations as the dependent variable, and construct the key independent variable as a dummy variable *AfternoonExp* to indicate that this observation is the afternoon elicitation. The results are reported in Table 6. Columns (1) and (4) report the whole-sample estimates for income and hours expectations, separately. The estimates suggest that overall, the afternoon expectations do not differ statistically from the morning ones for either income or hours, implying that expectations are on average sticky within a day.

However, it is still interesting to explore whether there is a heterogeneous effect underlying the overall estimates, especially for good and bad days, because the model suggests that if anything, the updating direction will be different on good and bad days.

We again separate the sample by whether the realized wage between the two elicitations is higher or lower than the morning expected wage. Column (2) reports the “bad-day” results for income. It shows that if the realized wage is below the morning expected wage, the afternoon income expectations are reduced significantly by 40.370 CNY (5.767 USD), which is about 4% of the sample mean of income. Column (3) shows the “good-day” results, in which wage realization is higher than the morning expectation, and the afternoon income expectations are increased significantly by 7.527 CNY (1.075 USD), which is about 0.8% of the sample mean. Interestingly, Columns (5) and (6) show no such significant adjustment exists in the hours expectations, regardless of whether it is a good or a bad day.

Overall, we find that income expectations do increase (decrease) significantly toward “rational” expectations, when the wage realization is higher (lower) than expected in the first half of the shift. The evidence of rational adjustment in income expectations is qualitatively consistent with Kőszegi and Rabin’s (2006) assumption. On the other hand, we find that expectations regarding daily working hours are sticky, limiting their effect on drivers’ responses to wage shocks.

[**Table 6 here**]

Finding 4. The adjustments of daily income expectations are qualitatively consistent with rational expectations for wages, but there is little adjustment of hours expectations within a day.

6. Conclusion

This paper presents the first field study that directly elicits cabdrivers’ expectations of income and working hours and uses them to estimate a model of reference-dependent labor supply. We show that the expectations elicitation itself does not significantly change drivers’ working patterns and that the elicited expectations data contain meaningful information about drivers’ expectations. This allows us to more precisely estimate how their expectations affect labor supply. It also yields valuable information on how they form and adjust their expectations.

We find that cabdrivers’ stopping decisions are affected by both income and hours targets, and the effect of the hours target (which is distinct from the neoclassical direct effect of hours on preferences) is the stronger of the two. With regard to income targets, only the afternoon target has a significant effect on drivers' stopping decisions, but both the morning and afternoon hours targets have significant effects. Splitting the sample according to whether the realized wage exceeds the expected wage reveals a heterogenous pattern that is powerful evidence for an expectations-based two-target model like Crawford and Meng's (2011). With regard to expectations, we find that income and hours targets are mainly affected by their recent historical average outcomes. Within shifts, hours expectations are very stable, while income expectations adjust in the direction suggested by the wage realizations in the first half of the shift.

Several issues are worthy of further discussion. First, the fact that the hours target is relatively stable, combined with the results in Table 4 that the effects of the hours target are stronger than those of the income target, suggest that the hours target is a more important factor than previously thought and that a two-target model improves on the one-target models used in most previous work. As previously discussed, the effect of hours targeting is different from the neoclassical effect of working hours.

Moreover, we find that drivers' elicited expectations explain their behavior somewhat better than would rational expectations proxied by historical outcomes as in Crawford and Meng (2011). Appendix Table A3 shows that R2 is slightly higher when we add elicited expectations, compared to that when we add historical outcomes. If we add the two sets of expectation measures in the regression simultaneously, the effect of elicited income expectations is not statistically different from that of historical average income, while the effect of elicited hours expectations is significantly larger than that of historical hours. In general, direct elicited expectations data have some value added relative to the historical average proxies used in the literature. In situations where expectations are unobservable, the use of historical average as a proxy variable is used as a substitute for elicitations in previous studies (Crawford and Meng, 2011; Agarwal et al., 2015). Our elicited expectations generally validate this proxy approach, but add some information.

Acknowledgments: Crawford acknowledges support from All Souls College and the University of Oxford; and Crawford and Meng acknowledge support from the European Research Council under grant ERC-2013-AdG 339179-BESTDECISION. Jin acknowledges the support of the National Natural Science

Foundation of China (grant number 72203225) and the Science Foundation of Ministry of Education of China (grant number 22YJC790050). Meng acknowledges the support of the National Natural Science Foundation of China (grant numbers 72394393, 72225002). Yao acknowledges the support of the National Natural Science Foundation of China (grant numbers 72394394, 72273077). The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The contents reflect only the authors' views and not the views of the funders, the European Research Council or the European Commission, and the European Union is not liable for any use that may be made of the information contained therein. Some exposition borrows from Crawford and Meng (2011) and Blow and Crawford (2023).

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman (2011): "Reference Points and Effort Provision," *American Economic Review*, 101, 470–492.

Agarwal, Sumit, Mi Diao, Jessica Pan, and Tien Foo Sing (2015): "Are Singaporean Cabdrivers Target Earners?" Available at SSRN 2338476.

Banerji, Abhijit, and Neha Gupta (2014): "Detection, Identification, and Estimation of Loss Aversion: Evidence from an Auction Experiment," *American Economic Journal: Microeconomics*, 6.1, 91-133.

Barberis, Nicholas, and Wei Xiong (2009): "What Drives the Disposition Effect: An Analysis of a Long-standing Preference-Based Explanation," *Journal of Finance*, 64, 751–784.

Ben-David, Itzhak, and David Hirshleifer (2012): "Are Investors Really Reluctant to Realize Their Losses? Trading Responses to Past Returns and the Disposition Effect," *Review of Financial Studies*, 25, 2485–2532.

Blow, Laura, and Vincent P. Crawford (2023): "Meaningful Theorems: Nonparametric Analysis of Reference-dependent Preferences," manuscript, https://papers.ssrn.com/sol3/papers.cfm?abstract_id =4645210.

Brandon, Alec, Colin F. Camerer, John A. List, Ian Muir, Jenny Wang (2023): "Evaluating the Evidence of Daily Income Targeting with Experimental and Observational Data," Working Paper.

Brown, Alexander L., Taisuke Imai, Ferdinand M. Vieider, and Colin F. Camerer (2021): "Meta-Analysis of Empirical Estimates of Loss-Aversion," Working paper.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler (1997): "Labor Supply of New York City Cabdrivers: One Day at a Time," *Quarterly Journal of Economics*, 112, 407-441.

Chang, Tom, and Tal Gross. (2014): "How Many Pears would a Pear Packer Pack if a Pear Packer could Pack Pears at Quasi-exogenously Varying Piece Rates?" *Journal of Economic Behavior and Organization* 99, 1-17.

Chou, Yuan K. (2002): "Testing Alternative Models of Labour Supply: Evidence from Taxi Drivers in Singapore," *The Singapore Economic Review*, 47.1, 17-47.

Crawford, Vincent P., and Meng, Juanjuan (2011): "New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 101, 1912–1932.

Dupas, Pascaline, Jonathan Robinson, and Santiago Saavedra (2020): "The Daily Grind: Cash Needs and Labor Supply," *Journal of Economic Behavior and Organization*, 177, 399-414.

Ericson, Keith M. Marzilla, and Andreas Fuster (2011): "Expectations as Endowmenets: Evidence on Reference-dependent Preferences from Exchange and Valuatiion Experiments," The *Quarterly Journal of Economics*, 126, 1879–1907.

Farber, Henry S. (2005): "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers," *Journal of Political Economy*, 113, 46–82.

Farber, Henry S. (2008): "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers," *American Economic Review*, 98, 1069-1082.

Farber, Henry S. (2015): "Why you Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers," *Quarterly Journal of Economics*, 130, 1975–2026.

Fehr, Ernst, and Lorenz Goette (2007): "Do Workers Work More if Wages are Higher? Evidence from a Randomized Field Experiment," *American Economic Review*, 97, 298-317.

Genesove, David and Christopher Mayer (2001): "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics*, 116, 1233-1260.

Gneezy, Uri, Lorenz Goette, Charles Sprenger, and Florian Zimmermann (2017): "The Limits of Expectations-Based Reference Dependence," *Journal of the European Economic Association*, 15, 861–876.

Goette, Lorenz, Thomas Graeber, Alexandre Kellogg, and Charles Sprenger (2020): "Heterogeneity of Gain-Loss Attitudes and Expectations-Based Reference Points," manuscript, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3589906.

Goldberg, Jessica (2016): "Kwacha Gonna Do? Experimental Evidence about Labor Supply in Rural Malawi," *American Economic Journal: Applied Economics*, 8.1, 129-49.

Greenwood, Robin, and Andrei Shleifer (2014): "Expectations of Returns and Expected Returns," *The Review of Financial Studies*, 27.3, 714-746.

Heffetz, Ori, and John A. List (2014): "Is the Endowment Effect an Expectations Effect?" *Journal of the European Economic Association*, 12,1396–1422.

Heffetz, Ori (2021): "Are Reference Points merely Lagged Beliefs over Probabilities?" *Journal of Economic Behavior and Organization*, 181, 252-269.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler (1990): "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 98, 1325-1348.

Kahneman, Daniel, and Amos Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263-292.

Karle, Heiko, Georg Kirchsteiger, and Martin Peitz (2015): "Loss Aversion and Consumption Choice: Theory and Experimental Evidence," *American Economic Journal: Microeconomics* ,7.2, sp101-20.

Kőszegi, Botond, and Matthew Rabin (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121, 1133-1165.

Kuchler, Theresa, and Basit Zafar (2019): "Personal Experiences and Expectations about Aggregate Outcomes," *The Journal of Finance*, 74.5, 2491-2542.

Li, Teng, Wenlan Qian, Wei Xiong, and Xin Zou (2022): "Employee Output Response to Stock Market Wealth Shocks," *Journal of Financial Economics*, 146.2, 779-796.

Lien, Jaimie W. and Jie Zheng, (2015) "Deciding When to Quit: Reference-Dependence over Slot Machine Outcomes", *American Economic Review: Papers and Proceedings*, 105, 366-370

Malmendier, Ulrike, and Stefan Nagel (2011): "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?" *The Quarterly Journal of Economics*, 126.1, 373-416.

Malmendier, Ulrike, and Stefan Nagel (2016): "Learning from Inflation Experiences," *The Quarterly Journal of Economics*, 131.1, 53-87.

Meng, Juanjuan and Xi Weng (2018): "Can Prospect Theory Explain the Disposition Effect? A New Perspective on Reference Points," *Management Science* 64, 3331-3351.

Odean, Terence (1998): "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance*, 53, 1775–1798.

Oettinger, Gerald S. (1999): "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors," *Journal of Political Economy*, 107, 360-392.

Pope, Devin G., and Maurice E. Schweitzer (2011): "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review*, 101, 129–57.

Post, Thierry, Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler (2008): "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show," *American Economic Review*, 98, 38-71.

Saia, Alessandro (2017): "Wishing for Bad Weather: Demand Shocks and Labor Supply of Bologna Pizza Vendors," Working paper.

Sheldon, Michael (2016): "Income Targeting and the Ridesharing Market," Unpublished manuscript.

Song, Changcheng (2016): "An Experiment on Reference Points and Expectations" Available at SSRN 2580852.

Sprenger, Charles (2015): “An Endowment Effect for Risk: Experimental Tests of Stochastic Reference Points” *Journal of Political Economy*, 123.6, 1456-1499.

Stafford, Tess M. (2015): “What do Rishermen Tell Us that Taxi Drivers Do Not? An Empirical Investigation of Labor Supply,” *Journal of Labor Economics*, 33.3, 683-710.

Thakral, Neil, and Linh T. Tô (2021): “Daily Labor Supply and Adaptive Reference Points,” *American Economic Review*, 111.8, 2417-2443.

Tversky, Amos, and Daniel Kahneman (1991): “Loss Aversion in Riskless Choice: A Reference-Dependent Model,” *Quarterly Journal of Economics*, 106, 1039-61.

Figures

Figure 1: Shift-level earnings (working hours) and expectations by day of the week

This figure presents the average shift-level earnings (working hours) and expectations, by day of the week.

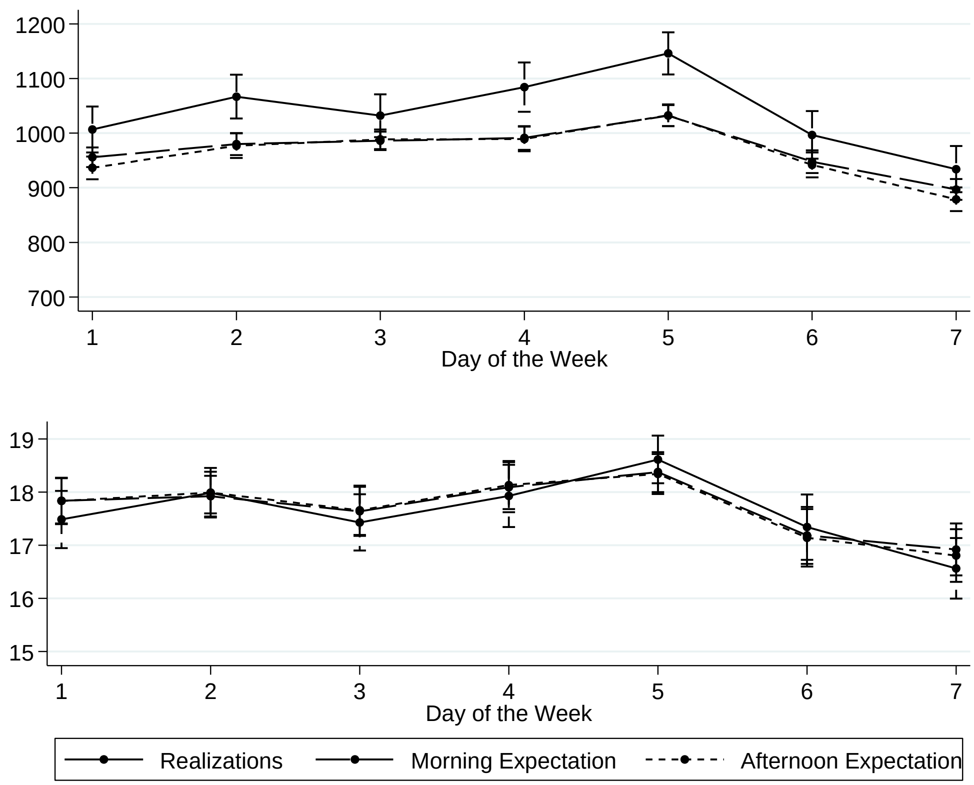


Figure 2: Shift-level earnings (working hours) and expectations by day of the week (for subsamples)

This figure presents the average shift-level earnings (working hours) and expectations for subsamples, by day of the week. Panel A restricts the sample to shifts with the realized wage between the two expectation elicitations lower than morning expected wage. Panel B restricts the sample to shifts with the realized wage between the two expectation elicitations higher than morning expected wage.

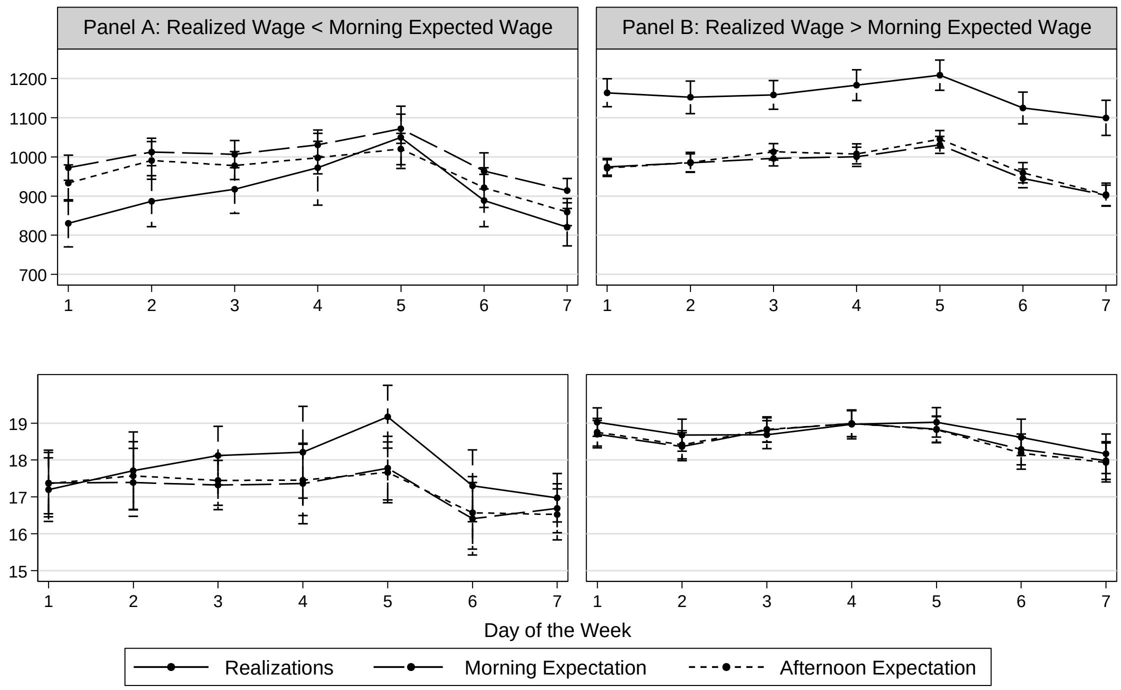
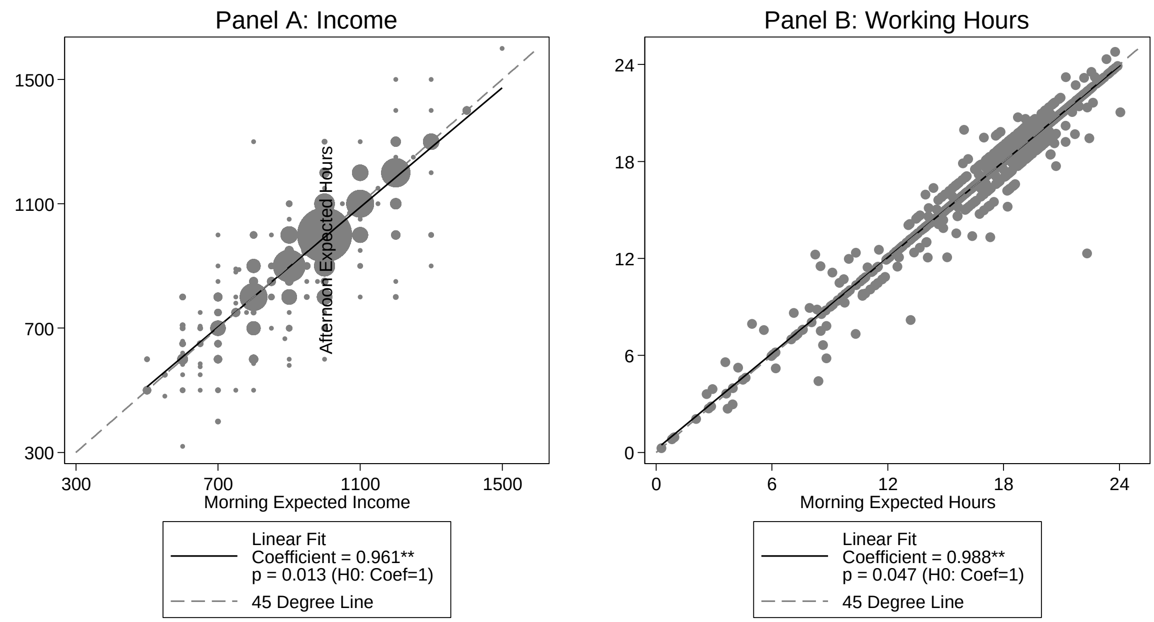


Figure 3: Morning and afternoon expectations

This figure presents the morning expectations against afternoon ones. Panel A presents the expectation of income, and Panel B presents the expectation of working hours.



Tables

Table 1: Summary statistics

This table reports summary statistics at the shift (Panel A), driver (Panel B), and trip (Panel C). *Wage* is the earnings per hour.

	Obs	Mean	S.D.	p25	p50	p75
Panel A: Shift level						
Income (CNY)	10137	972.2	366.4	759	1013	1228
Working hours	10137	16.5	5.1	14.5	17.9	19.9
Wage (CNY)	10137	61.0	28.2	51.4	58.9	66.5
Morning income expectation	1181	969.0	156.7	900.0	1000.0	1000.0
Afternoon income expectation	1181	962.3	172.2	890.0	1000.0	1050.0
Morning hours expectation	1181	17.7	3.6	16.5	18.4	20.0
Afternoon hours expectation	1181	17.7	3.6	16.4	18.4	20.0
Panel B: Driver level						
Monthly income (CNY)	131	10011.5	5164.9	7000	9000	12000
Monthly consumption (CNY)	131	5422.1	2994.5	3000	5000	7000
Monthly fee of the taxi (CNY)	121	5736.8	1414.9	5300	5400	5500
Panel C: Trip level						
Temperature (Degrees Celsius)	241761	27.2	4.0	24.5	27.5	30.0
Precipitation (Millimeters)	241761	0.1	0.6	0.0	0.0	0.0

Table 2: The effects of the intervention on stopping decision (treated vs. untreated shifts)

This table reports the effects of the experimental intervention on wage-hours elasticity at shift level, and stopping decision at trip level, by comparing the treated and untreated shifts of treated individuals during the experiment period. Time controls include day-of-week fixed effects, day-of-month fixed effects, month fixed effects, the clock hour of shift and trip ending time, holiday indicator, and the DiDi event indicators. Weather controls consist of temperate and precipitation together with their square terms. Other controls include the ratio of kilometers traveled with passengers, the number of Expectation Treatments in history, and the number of Neutral Treatments in history. Standard errors in parentheses are clustered at the individual level. *, ** and *** denote significance at the 10%, 5%, and 1% level, respectively.

Dependent variable:	(1) Log hours	(2) Log hours (IV)	(3) Stop
Neutral Treatment × Log wage	0.020	5.372	
	(0.240)	(31.120)	
Expectation Treatment × Log wage	0.219	2.115	
	(0.144)	(10.947)	
Log wage	-1.240***	41.321	
	(0.114)	(167.861)	
Neutral Treatment × Cumulative Income (/10000)			0.016
			(0.148)
Expectation Treatment × Cumulative Income (/10000)			0.039
			(0.092)
Cumulative Income (/10000)			0.215
			(0.144)
Neutral Treatment × Cumulative Hours (/100)			-0.025
			(0.089)
Expectation Treatment × Cumulative Hours (/100)			-0.054
			(0.062)
Cumulative Hours (/100)			1.254***
			(0.094)
Neutral Treatment	-0.033	-21.092	0.000
	(0.969)	(124.433)	(0.004)
Expectation Treatment	-0.839	-7.788	0.001
	(0.582)	(42.003)	(0.003)
Driver Fixed Effects	Yes	Yes	Yes
Time Controls	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes
Observations	10,137	10,137	241,761
R^2	0.505		0.162
F-stat first stage		0.036	
Number of Drivers	131	131	131
p value (H0: Neutral × Log wage = Expectation × Log wage)	0.462	0.896	
p value (H0: Neutral × Cum Income= Expectation × Cum Income)			0.888
p value (H0: Neutral × Cum Hours = Expectation × Cum Hours)			0.785

Table 3: The effects of intervention on income and working hours (treated vs. untreated shifts)

This table reports the effects of experimental intervention on income and working hours at shift level, by comparing the treated and untreated shifts of treated individuals during the experiment period. Time controls include day-of-week fixed effects, day-of-month fixed effects, month fixed effects, the clock hour of shift ending time, holiday indicator, and the DiDi event indicators. Weather controls consist of temperate and precipitation together with their square terms. Other controls include the ratio of kilometers traveled with passengers, the number of Expectation Treatments in history, and the number of Neutral Treatments in history. Standard errors in parentheses are clustered at the individual level. *, ** and *** denote significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)
Dependent variable:	Income	Hours
Omitted untreated		
Neutral Treatment	24.110**	0.500***
	(10.903)	(0.169)
Expectation Treatment	17.596**	0.437***
	(8.757)	(0.155)
Driver Fixed Effects	Yes	Yes
Time Controls	Yes	Yes
Weather Controls	Yes	Yes
Other Controls	Yes	Yes
Observations	10,137	10,137
R^2	0.424	0.441
Number of Drivers	131	131
p value (H0: Neutral Treatment = Expectation Treatment)	0.501	0.584

Table 4: The effects of elicited expectation on stopping decision

This table reports the effects of elicited expectation on stopping decision at trip level. Columns (1) to (3) use the full sample. Column (4) restricted the sample to shifts with the realized wage between the two expectation elicitations lower than the morning expected wage. Column (5) restricted the sample to shifts with the realized wage between the two expectation elicitations higher than the morning expected wage. Both columns (4) and (5) include only trips after the afternoon expectation elicitation. Time controls include day-of-week fixed effects, day-of-month fixed effects, month fixed effects, the clock hour of trip ending time, holiday indicator, and the DiDi event indicators. Weather controls consist of temperate and precipitation together with their square terms. Other controls include the ratio of kilometers traveled with passengers, the number of Expectation Treatments in history, and the number of Neutral Treatments in history. Standard errors in parentheses are clustered at the individual level. *, ** and *** denote significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Stopping decision				
Sample:	All			Trips after the Second elicitation	
				Early realized wage < Morning expected wage	Early realized wage > Morning expected wage
Cumulative Income/10000	-0.268	-0.161	-0.227	-0.553	0.505
	(0.177)	(0.176)	(0.179)	(1.017)	(0.306)
Cumulative Hours/100	1.132***	1.086***	1.056***	0.073	0.232
	(0.159)	(0.156)	(0.162)	(0.743)	(0.285)
1(Cumulative Income >Morning Expectation)	0.048***		0.017	0.007	0.015
	(0.010)		(0.011)	(0.041)	(0.012)
1(Cumulative Income >Afternoon Expectation)		0.051***	0.037***	0.066*	0.009
		(0.011)	(0.013)	(0.039)	(0.013)
1(Cumulative Hours > Morning Expectation)	0.216***		0.143***	0.223***	0.126***
	(0.021)		(0.028)	(0.040)	(0.040)
1(Cumulative Hours > Afternoon Expectation)		0.213***	0.093***	0.010	0.106***
		(0.021)	(0.026)	(0.041)	(0.037)
Driver Fixed Effects	Yes	Yes	Yes	Yes	Yes
Time Controls	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes
Observations	30,478	30,254	30,206	2,598	8,709
R^2	0.227	0.229	0.234	0.255	0.250
Number of Drivers	129	129	129	97	123
p value, H0:					
1(Cum. Inc.>Mor. Exp.) = 1(Cum. Hours>Mor. Exp.)	0.000		0.000	0.000	0.007
1(Cum. Inc.>Aft. Exp.) = 1(Cum. Hours>Aft. Exp.)		0.000	0.060	0.349	0.015
1(Cum. Inc.>Mor. Exp.) = 1(Cum. Inc.> Aft. Exp.)			0.324	0.412	0.783
1(Cum. Hours.>Mor. Exp.) =1(Cum. Hours>Aft. Exp.)			0.310	0.005	0.778

Table 5: The factors of elicited expectation

This table reports the factors of the elicited expectation. The sample is restricted to the expectation-treated shifts. For each shift, we focus on the morning expectation. In columns (1) and (2), the dependent variable is the expectation of income. In columns (3) and (4), the dependent variable is the expectation of working hours. In columns (1) and (3), we control for individual-level characteristics. In columns (2) and (4), we control for the individual fixed effects. We analyze the effects of past income and working hours for each driver and day of the week. *Historyincome1to5* (*Historyhour1to5*) is the average shift income (working hours) of the past five shifts for the same driver and day of the week. *Historyincome6to10* (*Historyhour6to10*) is the average shift income (working hours) of the past sixth to tenth shifts for the same driver and day of week. Time controls include day-of-week fixed effects, day-of-month fixed effects, month fixed effects, the clock hour of treatment time, holiday indicator, and the DiDi event indicators. Weather controls consist of temperate and precipitation together with their square terms. Other controls include the ratio of kilometers traveled with passengers, the number of Expectation Treatments in history, and the number of Neutral Treatments in history. Standard errors in parentheses are clustered at the individual level. *, ** and *** denote significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)
Dependent variable:	*Expected income*		*Expected hours*	
Historyincome1to5	0.183***	0.075***		
	(0.036)	(0.023)		
Historyincome6to10	0.042	0.016		
	(0.030)	(0.027)		
Historyhour1to5			0.418***	0.134**
			(0.053)	(0.059)
Historyhour6to10			0.113**	0.022
			(0.053)	(0.052)
Daily consumption	0.048		-0.000	
	(0.080)		(0.001)	
Daily fee of the taxi	-0.292		-0.005	
	(0.242)		(0.004)	
Driver Fixed Effects	No	Yes	No	Yes
Time Controls	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes
Observations	1,034	1,093	1,034	1,093
R^2	0.373	0.311	0.518	0.190
Number of Drivers	118	126	118	126
p value (H0: Historyincome1to5 = Historyincome6to10)	0.012	0.101		
p value (H0: Historyhour1to5 = Historyhour6to10)			0.001	0.249

Table 6: The adjustments of elicited expectation

This table reports the adjustments of elicited expectation. The sample is restricted to the expectation-treated shifts. For each shift, there are two observations: The morning and afternoon expectations, respectively. Therefore, the sample is at shift-expectation level. In columns (1) to (3), the dependent variable is the expectation of income. In columns (4) to (6), the dependent variable is the expectation of working hours. The main independent variable *AfternoonExp* is an indicator that equals one if the observation is the afternoon expectation and equals zero if otherwise. *BelowMorning (OverMorning)* restricts the sample to the shifts with the wage between the two elicitations is less (more) than the expected wage in the morning expectation. Time controls include day-of-week fixed effects, day-of-month fixed effects, month fixed effects, the clock hour of treatment time, holiday indicator, and the DiDi event indicators. Weather controls consist of temperate and precipitation together with their square terms. Other controls include the ratio of kilometers traveled with passengers, the number of Expectation Treatments in history, and the number of Neutral Treatments in history. Standard errors in parentheses are clustered at the individual level. *, ** and *** denote significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	*Expected income*			*Expected working hours*		
Sample	All	Below Morning	Over Morning	All	Below Morning	Over Morning
AfternoonExp	-4.736	-40.370***	7.527**	0.004	0.025	-0.002
	(3.193)	(7.394)	(3.189)	(0.023)	(0.055)	(0.025)
Driver Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,156	552	1,604	2,156	552	1,604
R^2	0.269	0.346	0.289	0.126	0.315	0.142
Number of Drivers	126	97	123	126	97	123

(For Online Publication)

Online Appendix of

"Expectations-based Reference-Dependence and Labor Supply: Eliciting Cabdrivers' Expectations in the Field"

By VINCENT P. CRAWFORD, MIAO JIN, JUANJUAN MENG, AND LAN YAO

Appendix A: Additional Figures and Tables

Figure A1: Share of treatments by day of the week

This figure presents the average share of treatments by day of the week for Expectation and Neutral Treatments, respectively. Specifically, for each driver, we calculate the share of treatments among the seven days of the week, and then take the average for all drivers.

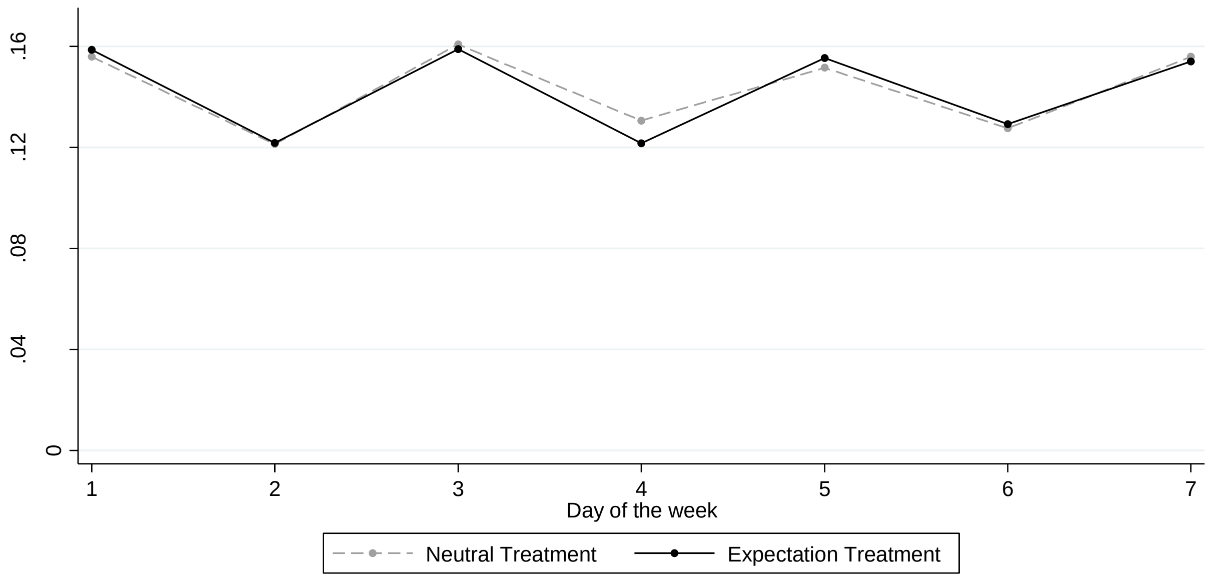


Table A1: Summary statistics of cabdrivers' background information

This table reports summary statistics of cabdrivers' background information.

	Obs	Mean	S.D.	p25	p50	p75
Marital status (Married, Widowed =1; Single, Divorced=0)	129	0.94	0.24	1	1	1
Family size	131	3.98	1.60	3	4	5
Number of children under 18	131	0.60	0.78	0	0	1
Number of employed family members	131	2.31	1.29	2	2	3
Monthly income (CNY)	131	10011.45	5164.91	7000	9000	12000
Monthly consumption (CNY)	131	5422.14	2994.49	3000	5000	7000
Daily lunch fee (CNY)	129	21.30	10.92	15	20	20
Monthly fee of the taxi (CNY)	121	5736.84	1414.85	5300	5400	5500
Monthly consumption categories						
Rental (CNY)	131	637.86	1216.90	0	0	1200
Mortgage (CNY)	131	1196.18	1991.06	0	0	2000
Children's' education (CNY)	121	1149.11	1592.65	0	1000	2000
Other debt (CNY)	131	3739.08	20999.02	0	0	0
Disease						
Cervical spondylopathy (=1)	131	0.53	0.50	0	1	1
Lumbar disc protrusion (=1)	131	0.40	0.49	0	0	1
Scapulohumeral periarthritis (=1)	131	0.41	0.49	0	0	1
Stomach trouble (=1)	131	0.53	0.50	0	1	1
Prostatitis (=1)	131	0.23	0.42	0	0	0
None of the above (=1)	131	0.18	0.38	0	0	0

Table A2: The effects of the intervention on stopping decision using local linear regression (treated vs. untreated shifts)

This table reports in each cell the *p*-value from a test of the differences in stopping decisions between the untreated and experimental treated samples using local linear regression. The rows correspond to different hours during the shift between 2 to 17 hours, roughly corresponding to the tenth and ninetieth percentile of the distribution of ending times of trips. For each hour, a local linear regression is conducted using trips ending within that. In each regression, the dependent variable is the drivers' stopping decisions, and the main independent variables are cumulative income earned during different hours of the shift and their interactions with the intervention indicators. For each regression, we jointly test the coefficients of the interaction terms between an intervention indicator and the income earned across all hours. The *p*-values from these joint tests are reported in this table. The columns correspond to the comparisons between two of the three samples: Untreated, Neutral Treatment, and Expectation Treatment. All regressions include full set of controls: intervention indicators, driver fixed effects, cumulative hours, time controls (day-of-week fixed effects, day-of-month fixed effects, month fixed effects, the clock hour of shift/trip ending time, holiday indicator, and the DiDi event indicators), weather controls (temperate and precipitation together with their square terms), other controls (the ratio of kilometers traveled with passengers, the number of Expectation Treatments in history, and the number of Neutral Treatments in history). Standard errors are clustered at the individual level.

	Untreated vs. NeutralTreatment	Untreated vs. ExpectationTreatment	NeutralTreatment vs. ExpectationTreatment
[2, 3)	0.279	0.150	0.541
[3, 4)	0.549	0.553	0.837
[4, 5)	0.914	0.415	0.694
[5, 6)	0.595	0.657	0.683
[6, 7)	0.722	0.267	0.750
[7, 8)	0.150	0.213	0.765
[8, 9)	0.529	0.238	0.827
[9, 10)	0.334	0.324	0.538
[10, 11)	0.565	0.692	0.826
[11, 12)	0.710	0.640	0.600
[12, 13)	0.397	0.432	0.274
[13, 14)	0.742	0.739	0.757
[14, 15)	0.728	0.860	0.331
[15, 16)	0.595	0.799	0.533
[16, 17)	0.139	0.496	0.230

Table A3: The effects of elicited expectation and historical outcomes on stopping decision

This table reports the effects of elicited expectation and historical outcomes on stopping decision at trip level. Columns (1) to (3) add morning elicited expectations, afternoon elicited expectations, and historical outcomes, respectively. Column (4) add morning elicited expectations and historical outcomes simultaneously, and Column (5) add afternoon elicited expectations and historical outcomes simultaneously. *Historyincome1to5* (*Historyhour1to5*) is the average shift income (working hours) of the past five shifts for the same driver and day of week. Time controls include day-of-week fixed effects, day-of-month fixed effects, month fixed effects, the clock hour of trip ending time, holiday indicator, and the DiDi event indicators. Weather controls consist of temperate and precipitation together with their square terms. Other controls include the ratio of kilometers traveled with passengers, the number of Expectation Treatments in history, and the number of Neutral Treatments in history. Standard errors in parentheses are clustered at the individual level. *, ** and *** denote significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)
Dependent variable:			Stopping decision		
Cumulative Income/10000	-0.268	-0.161	0.094	-0.216	-0.169
	(0.177)	(0.176)	(0.205)	(0.192)	(0.193)
Cumulative Hours/100	1.132***	1.086***	1.116***	0.809***	0.782***
	(0.159)	(0.156)	(0.157)	(0.168)	(0.163)
1(Cumulative Income > Historyincome1to5)			0.057***	0.042***	0.043***
			(0.011)	(0.011)	(0.011)
1(Cumulative Income >Morning Expectation)	0.048***			0.033***	
	(0.010)			(0.009)	
1(Cumulative Income >Afternoon Expectation)		0.051***			0.038***
		(0.011)			(0.010)
1(Cumulative Hours > Historyhour1to5)			0.111***	0.092***	0.094***
			(0.014)	(0.013)	(0.013)
1(Cumulative Hours > Morning Expectation)	0.216***			0.193***	
	(0.021)			(0.020)	
1(Cumulative Hours > Afternoon Expectation)		0.213***			0.189***
		(0.021)			(0.020)
Driver Fixed Effects	Yes	Yes	Yes	Yes	Yes
Time Controls	Yes	Yes	Yes	Yes	Yes
Weather Controls	Yes	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes
Observations	30,478	30,254	30,575	30,478	30,254
R^2	0.227	0.229	0.213	0.246	0.248
Number of Drivers	129	129	129	129	129
p value, H0:					
1(Cum. Inc.> History)= 1(Cum. Inc.> Mor. Exp.)				0.602	
1(Cum. Inc.> History)= 1(Cum. Inc.> Aft. Exp.)					0.752
1(Cum. Hours.> History.)= 1(Cum. Hours> Mor. Exp.)				0.000	
1(Cum. Hours.> History)= 1(Cum. Hours> Aft. Exp.)					0.000

Appendix B: Experimental Instructions

Survey

Background Information

1. What is your marital status?

A. Single, never married B. Married C. Divorced D. Widowed

2. How many family members do you have? (including you and your immediate family, such as your parents, your spouse and his/her parents, and your children.)

3. How many children under 18 years old do you have?

4. How many people are employed in your family? (including yourself)

5. What is your monthly household income?

6. What are your monthly household expenses?

What is your monthly household expenditure on the following categories?

Rental: _____ Mortgage: _____ Children's education: _____ Other debt: _____

7. How much do you spend on lunch on a working day?

8. What is your monthly fee for the cab? ____ What day of the month do you pay the fee? ____

9. Do you have the following disease?

A. Cervical spondylopathy B. Lumbar disc protrusion C. Scapulohumeral periarthritis D. Stomach trouble E. Prostatitis F. None of the above

Working Pattern

1. How much do you expect to earn on a normal day of the week? (not excluding the fee and fuel expense)

Monday: ____ Tuesday: ____ Wednesday: ____ Thursday: ____

Friday: ____ Saturday: ____ Sunday: ____

2. What is the start and end time of your work on a normal day of the week?

Monday: ____ to ____ Tuesday: ____ to ____ Wednesday: ____ to ____

Thursday: ____ to ____ Friday: ____ to ____ Saturday: ____ to ____

Sunday: ____ to ____

3. Please select one answer that most closely describes your working pattern.

A. I have an income target on each working day

B. I have a working hours target on each working day

C. I have both an income target and an hours target on each working day

D. I do not have an income target or hours target on each working day

4. If you have an income or hours target, what is the basis of your target?

A. Expectations on the basis of working experience B. Family expenses

C. Income and working status of other cabdrivers D. Others ____

5. If you have an income or hours target, how long do you set the target?

A. Every day B. Every week C. Every month D. Every year

6. If you have an unanticipated lump-sum income on a working day, what is the lowest income that will make you stop working earlier? Each horizontal line indicates a form of the unanticipated income. Please circle the number, above which you will stop working earlier. If you do not stop working earlier, please tick the "never stop working earlier".

Fuel cost subsidy

0 100 200 300 400 500 600 700 800 900 1000

never stop working earlier ☐

Unexpected income from unexpected long trip

0 100 200 300 400 500 600 700 800 900 1000

never stop working earlier ☐

WeChat transfer by this research team

0 100 200 300 400 500 600 700 800 900 1000

never stop working earlier ☐

Tips from passengers

0 100 200 300 400 500 600 700 800 900 1000 never stop working earlier ☐

Winning a prize

0 100 200 300 400 500 600 700 800 900 1000 never stop working earlier ☐

7. On what days of the week do you work in this week?

Monday: ___ Tuesday: ____ Wednesday: ____ Thursday: ___ Friday: ___ Saturday: ___ Sunday: ____

Preference Tests

1. A bag contains 10 balls, of which 5 are red and 5 are blue. One ball is to be drawn at random from the bag and the color of the drawn ball will determine your payment. In the following table, you are asked to make seven choices, one on each row. For each decision row, you will have to choose between two options: Option A and Option B. Which option do you prefer to get paid?

Decision	Option A	Option B	Your choice (A or B)
1	Win 20 CNY if a Red ball is drawn (50% probability), loss 5 CNY if a Blue ball is drawn (50% probability)	0	
2	Win 20 CNY if a Red ball is drawn (50% probability), loss 8 CNY if a Blue ball is drawn (50% probability)	0	
3	Win 20 CNY if a Red ball is drawn (50% probability), loss 12 CNY if a Blue ball is drawn (50% probability)	0	
4	Win 20 CNY if a Red ball is drawn (50% probability), loss 15 CNY if a Blue ball is drawn (50% probability)	0	
5	Win 20 CNY if a Red ball is drawn (50% probability), loss 18 CNY if a Blue ball is drawn (50% probability)	0	
6	Win 20 CNY if a Red ball is drawn (50% probability), loss 20 CNY if a Blue ball is drawn (50% probability)	0	
7	Win 20 CNY if a Red ball is drawn (50% probability), loss 25 CNY if a Blue ball is drawn (50% probability)	0	

2. You face the following pair of concurrent decisions. Both choices will be payoff-relevant, i.e., the gains and losses will be added to your overall payment. First, examine both decisions, then indicate your choices, by circling the corresponding letter.

Decision (1): The first bag contains 4 balls, of which 1 is red and 3 are blue. One ball is to be drawn at random from the bag and the color of the drawn ball will determine your payment. Which option do you prefer to get paid?

A. A sure gain of 4 CNY whatever the color of the drawn ball.

B. Gain 20 CNY if a red ball is drawn (25% probability), and gain 0 CNY if a blue ball is drawn (75% probability).

Decision (2): The second bag contains 4 balls, of which 1 is yellow and 3 are black. One ball is to be drawn at random from the bag and the color of the drawn ball will determine your payment. Which option do you prefer to get paid?

C. A sure loss of 15 CNY whatever the color of the drawn ball.

D. Lose 20 CNY if a black ball is drawn (75% probability), and lose 0 CNY if a yellow ball is drawn (25% probability).

3. In the following tables, you are asked to make several choices, one on each row. For each decision row, you will have to choose between payments to the earlier day (Option A) and payments to the later day (Option B). Which option do you prefer to get paid?

In the first table, please decide if you would like the payment of 100 CNY today or the payment of more than 100 CNY in 30 days. Please answer for each decision row.

Decision	Today (Option A)	In 30 days (Option B)	Your choice (A or B)
1	100 CNY	101 CNY	
2	100 CNY	104 CNY	
3	100 CNY	107 CNY	
4	100 CNY	110 CNY	
5	100 CNY	113 CNY	
6	100 CNY	116 CNY	
7	100 CNY	119 CNY	
8	100 CNY	122 CNY	
9	100 CNY	125 CNY	
10	100 CNY	128 CNY	

In the second table, please decide if you would like the payment of 100 CNY in 30 days or the payment of more than 100 CNY in 60 days. Please answer for each decision row.

Decision	In 30 days (Option A)	In 60 days (Option B)	Your choice (A or B)
11	100 CNY	101 CNY	
12	100 CNY	104 CNY	
13	100 CNY	107 CNY	
14	100 CNY	110 CNY	
15	100 CNY	113 CNY	
16	100 CNY	116 CNY	
17	100 CNY	119 CNY	
18	100 CNY	122 CNY	
19	100 CNY	125 CNY	
20	100 CNY	128 CNY	

Expectation Questions

The morning expectation question at 4 AM:

How much do you expect to earn today? What time do you expect to stop work today?

The afternoon expectation question at 6 PM:

According to your work so far, how much do you expect to earn today? What time do you expect to stop work today?

Neutral Questions

(1) What is the highest degree or level of education you have completed?

(2) What is your eating habit?

(3) What do you enjoy doing in your leisure time? Stay at home or get together with friends to party?

(4) How true is this of you? “I am happy-go-lucky.”

(5) How true is this of you? “I ‘squirm’ at plays or lectures.”

(6) How long do you plan for the future?

(7) How true is this of you? “Completing today’s tasks and doing other necessary work come before tonight’s play.”

(8) How true is this of you? “When I want to achieve something, I set goals and consider specific means for reaching those goals.”

(9) How true is this of you? “I feel that it’s more important to enjoy what you are doing than to get work done on time.”

(10) How true is this of you? “I keep working at difficult, uninteresting tasks if they will help me get ahead.”

(11) When would you prefer to have a free dinner at a fancy restaurant? On Saturday in 1 month or 2 months?

(12) Do you accept this gamble? Flip a coin, if it comes down heads, you get 110 CNY; if it comes down tails, you lose 100 CNY. [17]

[17] Questions (6) through (12) were loosely associated with risk preference, time preference, or goal-setting behavior. Our aim was to explore if any aspects of these dimensions could impact behavior via the priming effect. For this purpose, we compared the drivers' behavior upon being asked these specific questions with their responses to other questions, using the analyses provided in the Preliminary Statistical Tests section. However, we didn't observe any significant differences between the two groups, indicating that these questions did not uniquely affect work behaviors.