Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment

By ERNST FEHR AND LORENZ GOETTE*

Most previous studies on intertemporal labor supply found very small or insignificant substitution effects. It is possible that these results are due to constraints on workers' labor supply choices. We conducted a field experiment in a setting in which workers were free to choose hours worked and effort per hour. We document a large positive elasticity of overall labor supply and an even larger elasticity of hours, which implies that the elasticity of effort per hour is negative. We examine two candidate models to explain these findings: a modified neoclassical model with preference spillovers across periods, and a model with reference dependent, loss-averse preferences. With the help of a further experiment, we can show that only loss-averse individuals exhibit a negative effort response to the wage increase. (*JEL* J22, J31)

The intertemporal substitution of labor supply has far-reaching implications for the interpretation of important phenomena. If, for example, the intertemporal substitution of labor supply is high, one may interpret the large variations in employment during business cycles as voluntary choices by the workers rather than involuntary layoffs. Intertemporal substitution also plays a crucial role in the propagation of shocks across periods (David Romer 1996; Robert G. King and Sergio Rebelo 1999). Previous studies have found little evidence for intertemporal substitution of labor. The estimated elasticities are often small and statistically insignificant, and sometimes even negative (see, e.g., N. Gregory Mankiw, Julio Rotemberg, and Lawrence Summers 1985; John Pencavel 1986; Joseph Altonji 1986; Richard Blundell 1994; David Card 1994; Blundell and Thomas E. MaCurdy 1999).[1]

The low estimates of intertemporal substitution are difficult to interpret, however, because of serious limitations in the available data. The life-cycle model of labor supply predicts intertemporal substitution with regard to *transitory* wage changes or wage changes the workers *anticipate.* Yet, the typical wage changes are not transitory; hence, they are associated with significant income effects. In addition, it seems almost impossible to infer reliably from existing data whether the workers anticipated the wage change. Furthermore, serious endogeneity problems arise, as both supply and demand conditions determine wages.[2] Thus, the typically available data require many auxiliary assumptions when testing the life-cycle model of labor supply.

Another issue arises if labor markets are characterized by a significant amount of job rationing

* Fehr: University of Zurich, Institute for Empirical Economic Research, Blümlisalpstrasse 10, CH-8006 Zurich (e-mail: efehr@iew.unizh.ch); Goette: University of Zurich, Institute for Empirical Economic Research, Blümlisalpstrasse 10, CH-8006 Zurich (e-mail: lorenz@iew.unizh.ch). This paper is part of the research priority program on the foundations of human social behavior funded by the University of Zurich. The authors also acknowledge support from the Swiss National Science Foundation under project number 101312–103898/1. This paper greatly benefited from the comments of two excellent referees. In addition, we thank George Akerlof, Henry Farber, David Huffman, Reto Jegen, Rafael Lalive, George Loewenstein, Jennifer Lerner, Stephan Meier, Matthew Rabin, Jason Riis, Alois Stutzer, Richard Thaler, and George Wu for their helpful comments.

[1] After reviewing a sizeable part of the literature, Card (1994) concludes, for instance, that the "very small magnitude of the estimated intertemporal substitution elasticities" can account for only a tiny fraction of the large person-specific, year-to-year changes in labor supply.

[2] Gerald Oettinger (1999) shows that if one neglects the endogeneity of wage changes, estimates of labor supply elasticities are severely downward-biased.

or other constraints on workers' labor supply. In fact, there is strong evidence suggesting that workers are not free to set their working hours (John C. Ham 1982; Shulamit Kahn and Kevin Lang 1991; William T. Dickens and Shelly Lundberg 1993), rendering the identification of the source of small intertemporal substitution effects difficult, even if the problems mentioned above could be solved. A small intertemporal substitution effect could be due to these constraints, or it could be that the behavioral assumptions behind the life-cycle model are wrong. Indeed, Colin F. Camerer et al. (1997) put forward the view that New York City cab drivers' daily labor supply is driven by nonstandard, reference dependent preferences that exhibit loss aversion around a target income level. This view has recently been called into question by Henry S. Farber (2004, 2005).

In this paper, we use an ideal dataset to study workers' responses to transitory wage changes. We conducted a randomized field experiment at a bicycle messenger service in Zurich, Switzerland. The bicycle messengers receive no fixed-pay component and are paid solely on commission. We have precise information for all the workers on the number of shifts they work and the revenues they generate per shift. A shift always comprises five hours, and workers in our sample worked at most one shift per day. A key feature of our experiment is the implementation of an *exogenous* and *transitory* increase of 25 percent in the commission rate. Therefore, we can be sure that unobserved supply or demand variations did not induce the change in the commission rate (i.e., the "wage" change). Each participant in the experiment knew ex ante the precise duration and size of the wage increase. Since the wage was increased only during four weeks, its impact on the workers' lifetime wealth is negligible.

In the firm under study, the messengers can freely choose how many shifts (hours) they work and how much effort they exert (to generate revenues). This means that our setting also provides an ideal environment for studying the behavioral foundations of labor supply. In our context, the absence of intertemporal substitution effects cannot be attributed to institutional constraints on labor supply. The exogenous change in the commission rate raises the returns from both the number of shifts and effort per shift. In contrast to earlier studies (Oettinger 1999; Camerer et al. 1997; Yuan K. Chou 2002), we have the unique opportunity of studying how hours *and* effort respond to the wage increase *and* how *overall* labor supply (i.e., the number of hours times the effort per hour) is affected.

Our experimental results show that the wage increase caused a large increase in overall labor supply. Our estimate of the intertemporal elasticity of substitution with regard to overall labor supply is between 1.12 and 1.25. This large effect is exclusively driven by the increase in the number of hours worked. In fact, the elasticity of hours worked with regard to the wage is higher than the elasticity of overall labor supply. The elasticity of hours is between 1.34 and 1.50, considerably higher than that found in previous studies. For example, Oettinger (1999) investigates how stadium vendors adjust their labor supply to changes in expected wages. He uses a set of ex ante predictors of game attendance, which are strongly related to the hourly wages of stadium vendors. His estimated elasticities range from 0.53 to 0.64.

The fact that the elasticity of hours (shifts) worked is larger than the overall labor supply elasticity suggests that the effort per hour decreased in response to the wage increase. And indeed, a detailed analysis indicates that effort per shift decreased by roughly 6 percent in response to the wage increase, which implies a wage elasticity of effort per shift of -0.24. These results confirm the nonexperimental evidence in previous studies of intertemporal labor substitution based on samples where workers were largely unconstrained in choosing hours and effort. Camerer et al. (1997) and Chou (2002) examined how cabdrivers, after having decided to work on a given day, vary their daily working time (which is a good proxy for daily effort) in response to wage variations. Both studies report that workers work fewer hours (provide less effort) on high-wage days, indicating a negative effort elasticity. Interpreting this evidence is difficult, however, as pointed out by Goette, David Huffman, and Fehr (2004) and Farber (2004, 2005). One problem is that the source of the variation in cabdrivers' wages is not completely clear. If, for example, there are common supply-side shocks (e.g., most drivers prefer not working on the Fourth of

July), then the supply of cabdriver hours will be small on these days and the ensuing wage will be high. As a result, there will be a negative correlation between wages and hours, although all individuals have neoclassical time-separable preferences. A second concern is a possible selection effect: higher wages may induce cabdrivers to work a few hours on days when they otherwise would not have worked. Such an effect may generate a negative correlation between daily wages and daily hours, even though all individuals behave exactly as the standard model predicts. Our results, however, are immune to both criticisms; that is, the reduction in effort observed in our data questions the standard neoclassical model with time-separable preferences. After all, the rise in the commission rate provides strong economic incentives for working more hours *and* for working harder during those hours.

We provide two reasonable extensions of the standard model that can, in principle, explain a negative effort elasticity. In the theory part of our paper, we show that a neoclassical model, in which last period's effort raises this period's marginal disutility of effort, is consistent with our evidence—workers who work in more periods may rationally decide to reduce effort per period. We also show that a rational choice model, with reference dependent preferences exhibiting loss aversion around the reference point (Goette, Huffman, and Fehr 2004), is also able to explain the evidence. The intuition behind this model is that workers with loss-averse preferences have a daily reference income level.[3] Daily incomes below the reference level are experienced as a "loss" and the marginal utility of income is large in the loss domain. In contrast, the marginal utility of income at and above the reference level decreases discontinuously to a lower level. Workers who temporarily earn higher wages are more likely to exceed the reference income level, hence, reducing their marginal utility of income and ultimately inducing them to provide less effort per shift. At the same time, however, workers with higher wages have a higher overall utility from working a shift, so that they can more easily cover the fixed costs of getting to work. Hence, they are likely to work more shifts.

There are thus two competing theories which are consistent with the facts. In order to discriminate between the two theories, we conducted another experiment based on the idea that loss aversion is a personality trait which affects behavior across several domains (Daniel Kahneman and Amos Tversky 2000; Simon Gaechter, Andreas Herrmann, and Eric Johnson 2005). In this experiment, we measured the individual worker's loss aversion in lottery choices. We then used these measures to examine whether the negative response of effort per shift is due to the existence of loss-averse workers. We indeed find that the degree of a worker's loss aversion contributes significantly to the negative effort elasticity. Moreover, it turns out that workers who do not show loss aversion in the lottery choices also do not have a significantly negative elasticity. Only workers with loss aversion reduce effort per shift significantly when paid a high wage.

Thus, the result of our second experiment favors the model with reference dependent preferences over the neoclassical model with "disutility spillovers" across periods. Of course, the evidence from the second experiment is not the ultimate arbitrator, but it suggests that future work should not disregard the loss aversion model because it could contribute to a deeper understanding of effort choices. At the same time, we should also point out that one-third of the workers in our sample did not exhibit loss aversion and a negative effort elasticity. Thus, future work should take the possibility of heterogeneous preferences more seriously. In addition, the results of our first experiment unambiguously show that whatever behavioral forces worked against the intertemporal substitution of labor, they were apparently not capable of generating a negative elasticity of the overall labor supply. The behavioral forces that worked in favor of intertemporal substitution outweighed any opposing forces.

The remainder of this paper is structured as follows. Section I describes the institutional environment and the details of the field experiment. Section II discusses the implications of different models of labor supply. Section III reports the results from the field experiment.

[3] Chip Heath, Richard Larrick, and George Wu (1999) provide evidence that goals often serve the function of a reference point.

We first report the impact of the wage increase on overall labor supply and then discuss how shifts responded. Finally, we present the evidence on how the wage increase affected the effort per shift. This section also describes the follow-up experiment and discusses the link between individual loss aversion and workers' effort responses. Section IV concludes the paper.

I. Experimental Setup

Our study is based on the delivery records of two Swiss messenger services, Veloblitz and Flash Delivery Services (henceforth "Flash"), which are located in Zurich. Each firm employs between 50 and 60 bicycle messengers. The available records contain information about when a messenger worked a shift, all deliveries he conducted during a shift, and the price of each delivery. Thus, we know which messengers worked a shift and how much revenue was generated during the shift for each day in the observation period. We first describe the organization of work at a bicycle messenger service and then present our experiment in more detail.

A. *Work at a Messenger Service*

Unless pointed out below explicitly, the arrangements are the same for the two messenger services, Veloblitz and Flash. When a potential worker applies for a job with one of the messenger services, an experienced messenger evaluates him or her with respect to fitness, knowledge of locations, names of streets, courtesy, and skill handling the CB radio. Once accepted as an employee, messengers can freely choose how many five-hour shifts they will work during a week. There are about 30 shifts available at Veloblitz and about 22 at Flash on each workday (Monday to Friday). In principle, messengers could work more than one shift per day, but none of them chose to do so during the experiment or in the months prior to the experiment. The shifts are displayed on a shift plan for every calendar week at the messenger service's office. There are two types of shifts, called "fixed" and "sign up." A "sign-up" shift simply means that a shift is vacant at a particular time. Any messenger can sign up to work that shift (e.g., on Wednesday from 8 a.m. to 1 p.m.). If a messenger commits to a "fixed" shift, he has to work that shift every week. For example, if a messenger chooses Wednesday from 8 a.m. to 1 p.m. as a fixed shift, he will have to fill that shift every Wednesday for at least six months. Thus, fixed shifts represent a commitment of several months and can be cancelled only with at least four weeks notice. Roughly two-thirds of the shifts are fixed. It is also important to note that the number and the allocation of fixed shifts across messengers remained the same during the entire experiment. The company refused to change the fixed shifts just because of the experiment. All shifts that are not fixed are available to any messenger. All workers participating in our study worked both fixed and variable shifts.

Two further items are worth mentioning. First, there is no minimum number of shifts that the messengers have to work at either messenger service. Second, both messenger services found filling the available shifts difficult. There is almost always at least one unfilled shift and, on average, almost three shifts per day remain unfilled. For example, during the period before the experiment, from September 1999 to August 2000, approximately 60 shifts remained unfilled every month. This implies that messengers are unlikely to be rationed in the choice of shifts.

Messengers receive no fixed wage. Their earnings are given *solely* as a fixed percentage w of their daily revenues. Hence, if a messenger carries out deliveries that generate revenues r during his shift, his earnings on that day will be wr. An important feature of the work environment concerns the fact that messengers have substantial discretion about how much effort to provide during a shift. They stay in contact with the dispatcher at the messenger service office only through CB radio. In order to assign a delivery, say, from location A to location B, the dispatcher will contact the messenger whom he thinks is closest to A to pick up the delivery. All messengers can listen in on the radio. If they believe that they are closer to A than the messenger originally contacted, they can get back to the dispatcher and say so and will then be assigned that delivery. Conversely, if the messenger does not want to carry out the delivery from location A to location B, he may not respond to the call. Messengers have, therefore, several means of increasing the number of deliveries they complete. They can ride at higher speed, follow the radio more actively, or find the shortest possible way to carry out a delivery.

Thus, work at a bicycle messenger service closely approximates a model where individuals are unconstrained in choosing how many shifts (hours) to work and how hard to work (i.e., how many deliveries to complete during a shift).

B. *The Experimental Design*

In order to evaluate the labor supply effect of a temporary wage increase, we randomly assigned those Veloblitz messengers who were willing to participate in the experiment to a treatment and a control group, and we implemented a fully anticipated temporary increase in the commission rate by roughly 25 percent for the treatment group. The commission rate for men in the treatment group was temporarily increased from $w = 0.39$ to $w = 0.49$ and the rate for women was temporarily increased from $w = 0.44$ to $w = 0.54$. The additional earnings for the messengers were financed by the Swiss National Science Foundation.

In order to participate in the experiment, all messengers had to complete a questionnaire at the beginning and end of each experimental period. The messengers were informed that a failure to complete all questionnaires meant they would not receive the additional earnings from the experiment. All messengers who finished the first questionnaire also filled in the remaining questionnaires.[4] Thus, the group of messengers who participated in the experiment was constant during the entire experiment, i.e., there was no attrition. Randomization into a treatment and a control group was achieved by randomly allocating the *participating* messengers into a group A and a group B. The randomization was based on the administrative codes that the messenger service uses to identify a messenger in its accounting system. All messengers at Veloblitz were assigned a number depending on the date when they started working for the company. The first messenger who worked at Veloblitz was assigned the number 1, the second 2, and so forth. The participating messengers with odd numbers were assigned to group A and participating messengers with even numbers to group B.

The messengers did not know that the purpose of the experiment was the study of labor supply behavior, nor did they realize that we received the full (anonymous) records of each messenger containing the number of shifts and the number of deliveries completed. If pressed, we told the participants that we wanted to study the relation between wages and job satisfaction. The purpose of our study was credible because the questionnaires contained several questions related to job satisfaction.[5]

For group A, we implemented a 25-percent increase in the commission rate during the four weeks in September 2000. The messengers in group B were paid their normal commission rate during this time period so that they could be used as a control group. In contrast, only the individuals in group B received a 25-percent increase in the commission rate during the four weeks in November 2000, while the members of group A received their normal commission rate and therefore served as a control group. Thus, a key feature of our experiment is that there were two experimental periods that lasted for four weeks and both group A and group B served as a treatment and a control group in one of the two experimental periods. This feature, in combination with our participation rule, implies that our design is perfectly balanced during the two treatment periods. Therefore, the point estimate of the treatment effect is completely independent of individual heterogeneity between our subjects. We will include messenger fixed effects in most of the analysis, however, to reduce the estimated standard errors.

[4] The messengers at Veloblitz who did not participate in the experiment were almost exclusively workers who were already quite detached from the company or who were on probationary shifts. The "detached" workers typically worked roughly one shift per week during the experiment and the months prior to the experiment.

[5] These features of the experiment ensure that our results cannot be affected by the Hawthorne effect. The Hawthorne effect means that subjects behave differently just because they know that the experimenters observe their behavior. Yet, our subjects did not know that we could observe their behavior during the wage increase. Moreover, since both the treatment group and the control group are part of the overall experiment, and since our key results rely on the comparison between these groups, we control for a potential Hawthorne effect.

Our experiment thus enables us to provide a very clean isolation of the impact of the temporary wage increase. If, for example, the implemented wage change increases labor supply, we should observe this increase both in the first and the second experimental period. In the first experimental period, the members of group A (who receive the higher wage in this period) should exhibit a larger labor supply than the members of group B, while the reverse should be true in the second experimental period—members of group B (who receive the higher wage in this period) should supply more labor.

Our experimental design also enables us to control for the income effect of the wage increase, i.e., we can identify the pure substitution effect for the participating messengers. We announced the experiment in the last week of August 2000 and all additional earnings from the experiment—regardless of whether subjects were members of group A or group B—were paid out after the end of the second experimental period in December 2000.[6] Thus, the budget constraint for both groups of participating messengers was affected in the same way. Due to the randomization of the participating messengers into groups A and B, the income effect cancels out if we identify the treatment effect by comparing the labor supply of the control and treatment groups.

As demand for delivery services varies from day to day and from month to month, it is useful to control for time effects. The available information about Flash enables us to identify possible time effects across treatment periods because both Veloblitz and Flash operate in the same market. There is a strong correlation between the total daily revenues at Veloblitz and Flash. When we compute the raw correlation between total revenues at the two firms over the two experimental periods plus the four weeks prior to the experiment, we find a correlation of 0.56 (Breusch-Pagan $\chi^2(1) = 18.93$, $p < 0.01$, $N = 60$ days). Even after removing daily effects from both series, the correlation is still 0.46 (Breusch-Pagan $\chi^2(1) = 13.16$, $p < 0.01$, $N = 60$ days). This shows that the revenues at the two firms are highly correlated, even over such a short time horizon.[7]

We believe that our experiment represents a useful innovation to the existing literature for several reasons. First, it implements a fully anticipated, temporary, and exogenous variation in the (output-based) wage rates of the messengers, which is key for studying the intertemporal substitution of labor. The experimental wage increase was large and provides a clear incentive for increasing labor supply. Moreover, the participating messengers are experienced, and daily fluctuations in their earnings are common. Hence, we experimentally implement a wage change in an otherwise familiar environment. Second, the data we obtained from Veloblitz allow us to study two dimensions of labor supply: hours as measured by the number of shifts, and effort as measured by the revenues generated per shift or the number of deliveries per shift. No other study that we are aware of can look at these two dimensions simultaneously. Third, we can combine the dataset with the full records from a second messenger service operating in the same market. This will prove useful for investigating any effect that the experiment might have had on the nonparticipating messengers at Veloblitz, and helps to control for demand variations over time.

II. Predictions

In this subsection, we derive predictions about labor supply behavior in our experiment. We use two types of models: neoclassical models and a model of reference dependent utility with loss-averse workers. In view of our results, we are particularly interested in the question of which kind of model is capable of predicting an increase in shifts (hours) worked *and* a decrease in effort per shift.

[6] In the time period between the announcement of the experiment and the beginning of the first treatment period, no new regular workers arrived at Veloblitz. Only workers who worked on probationary shifts arrived during this time period, and they were not allowed to participate in the experiment because they often leave the firm after a short time and lack the necessary skills. Including them in the experiment would have created the risk of attrition bias.

[7] If we add the eight months prior to the experiment, we find a correlation of about 0.75. To check the robustness of our results, we also include—in some of our regressions—the nonparticipating messengers at Veloblitz in the nonexperimental comparison group that is used to identify time effects.

A. *Neoclassical Model with Time-Separable Utility*

In this subsection, we integrate the institutional setting at our messenger service into a canonical model of intertemporal utility maximization with time-separable utility. We define the relevant time period to be one day. Consider an individual who maximizes lifetime utility

$$(1) \qquad U_o = \sum_{t=0}^{T} \delta^t u(c_t, e_t, x_t),$$

where $\delta < 1$ denotes the discount factor, $u(\ )$ represents the one-period utility function, c_t denotes consumption, e_t is effort in period t, and x_t denotes a variable that affects the preference for working on particular days. For example, a student who works a few shifts per week at Veloblitz may have higher opportunity costs for working on Fridays because he attends important lectures on Fridays. The utility function obeys $u_c > 0$, $u_e < 0$ and is strictly concave in c_t and e_t. The lifetime budget constraint for the individual is given by

$$(2) \qquad \sum_{t=0}^{T} \hat{p}_t c_t (1+r)^{-t} = \sum_{t=0}^{T} (\hat{w}_t e_t + y_t)(1+r)^{-t},$$

where $\hat{p}_t$ denotes the price of the consumption good, $\hat{w}_t$ the period, t wage per unit of e_t, and y_t nonlabor income. For convenience we assume that the interest rate r is constant and there is no uncertainty regarding the time path of prices and wages. The sign of the comparative static predictions is not affected by these simplifying assumptions.

In an Appendix available online,[8] we show that along the optimal path, the within-period decisions of a rational individual maximizing a time-separable concave utility function like (1), subject to constraint (2), can be equivalently represented in terms of the maximization of a static one-period utility function that is linear in income.[9] This static utility function can be written as

$$(3) \qquad v(e_t, x_t) = \lambda w_t e_t - g(e_t, x_t),$$

where $g(e_t, x_t)$ is strictly convex in e_t, and measures the discounted disutility of effort, x_t captures exogenous shifts in the disutility of effort, λ measures the marginal utility of life-time wealth, and w_t represents the discounted wage in period t. Thus, $\lambda w_t e_t$ can be interpreted as the discounted utility of income arising from effort in period t.[10]

Workers who choose effort according to (3) respond to an anticipated temporary increase in w_t with a higher effort e_t. A rise in w_t increases the marginal utility returns of effort, λw_t, which increases the effort level e_t^* that maximizes $v(e_t, x_t)$. The situation is a bit more complicated in our experiment, however, because the messengers can choose the number of shifts and the effort during the shift. Theoretically, the existence of shifts can be captured by the existence of a minimal effort level $\tilde{e}$ that has to be met by the worker or by the existence of fixed costs of working a shift. Intuitively, if there is a fixed cost of working a shift, an employee will work on a given day only if the utility of e_t^*, $v(e_t^*, x_t)$ is higher than the utility of not going to work at

[8] The Appendix is available at www.e-aer.org/data/mar07/20020849_data.zip.

[9] Our characterization is inspired by the results in Martin Browning, Angus Deaton, and Margaret Irish (1985) who show that the within-period decisions can be characterized in terms of the maximization of a static profit function.

[10] λ is constant along the optimal path of c_t and e_t. This has the important consequence that an *anticipated* temporary wage variation does not affect the marginal utility of lifetime wealth. Thus, anticipated temporary variations in wages (or prices) have no income effects. Yet, if there is a nonanticipated temporary increase in the wage, λ changes immediately after the new information about the wage increase becomes available, and remains constant at this changed level afterward. For our experiment, this means that the income effect stemming from the temporary wage increase has to occur immediately after the announcement of the experiment in August 2000. Thereafter, the marginal utility of lifetime wealth again remains constant so that there are no further changes in λ during the experiment. The difference in behavior between the treatment group and the control group during the two treatments can thus not be due to changes in λ. Note also that (3) not only describes the optimal effort choice in period t, but also is based on the optimal consumption decision in period t. For any change in effort, the consumption decision also changes in an optimal manner (see Appendix).

all. As a wage increase raises $v(e_t^*, x_t)$, workers are more likely to work on a given day, i.e., the number of shifts worked will increase.[11]

B. *Neoclassical Model with Nonseparable Utility*

The prediction of the previous subsection is, however, not robust to the introduction of nonseparable utility functions. To illustrate this, consider a simple example where

$$(4) \quad v(e_t, e_{t-1}) = \lambda e_t w - g(e_t(1 + \alpha e_{t-1})).$$

This example captures the intuition that if a worker worked yesterday, he has higher marginal cost of effort today. We assume, for simplicity, that $e_0 = 0$, that there are only two further time periods (period 1 and period 2), and that the wage is constant across time. If we ignore discounting, the two-period utility is given by $U = v(e_1, 0) + v(e_2, e_1)$. Therefore, if the wage is high enough to induce the worker to go to work in both periods, the worker chooses effort e_1^{**} and e_2^{**} according to

$$(5) \quad \lambda w = g'(e_1) + \alpha e_2 g'(e_2(1 + \alpha e_1));$$

$$(6) \quad \lambda w = g'(e_2(1 + \alpha e_1))(1 + \alpha e_1).$$

If work is supplied in both periods, an increase in e_1 causes a higher disutility of labor in period 2, which lowers e_2. Of course, rational workers take this effect into account when they decide on e_1, which means that the overall marginal disutility of e_1 is higher if e_2 is positive compared to when it is zero. In particular, if wages are low enough so that it is no longer worthwhile to work in period 2 ($e_2 = 0$), the first-order conditions are given by

$$(5') \qquad \lambda w = g'(e_1);$$

$$(6') \qquad \lambda w < g'(0)(1 + \alpha e_1).$$

A comparison of conditions (5) and (6) with conditions (5′) and (6′) shows that it is possible that the optimal effort e_1 according to (5′) is higher than e_1^{**} and e_2^{**}. In the online Appendix, we provide an explicit example that proves this point. This possibility arises because the marginal disutility of working in each of the two periods, which is indicated by the right-hand side of (5) and (6), is higher than the marginal disutility of working only in period 1, which is given by $g'(e_1)$. In the context of our experiment, this means that messengers who work more shifts when the wage is high may rationally decide to reduce the effort per shift.

The simple model above does not predict that workers who work more shifts (days) will necessarily reduce their effort per shift. It allows for only this possibility. If the wage increase is large enough, it is also possible that workers who behave according to this model raise their effort per shift. There is, however, one prediction that follows unambiguously from a neoclassical approach regardless of whether utility is time separable or not. Browning, Deaton, and Irish (1985) have shown that a general neoclassical model predicts that overall labor supply, $\sum e_t$, increases in high-wage periods in response to a temporary increase in wages. Applied to our context, this means that during the four-week period where the wage is higher for the treatment group, the total revenue (or the total number of deliveries) of the treatment group should exceed the total revenue (or the total number of deliveries) of the control group.

C. *Reference Dependent Utility*

Another potential explanation for why effort per shift might decrease in response to a temporary wage increase is that individuals could have preferences that include a daily income target $\tilde{y}$ that serves as a reference point. The crucial element in this approach is that if a person falls short of his or her target, he or she is assumed to experience an additional psychological cost, which is not present if income varies above the reference point. This explanation is suggested by the large number of studies indicating reference dependent behavior (for a selection of papers on this see Kahneman and Tversky 2000). Evidence from psychology (Heath, Larrick, and Wu 1999) suggests that the marginal utility of a dollar below the target is strictly higher than

[11] More formally, the wage increase raises the utility of going to work for all x. Hence, the participation condition will be met for more states x.

the marginal utility of a dollar above the target.[12] A *daily* income target seems plausible for bike messengers in our sample because their daily incomes are a salient feature of their work environment. The messengers keep receipts from each delivery made on a shift. This makes them acutely aware of how much money they earn from each completed delivery. The messengers also turn in the receipts at the end of the shift, making it difficult for them to keep track of how much money they earned over several shifts. A daily income target may also serve the messengers as a commitment device for the provision of effort during the shift. Zurich is rather hilly and riding up the hills several times during a shift requires quite some effort—in particular if the weather is bad or toward the end of a shift. A daily income target may thus help the messengers overcome a natural tendency to "shirk" that arises from a high marginal disutility of effort.

As in Goette, Huffman, and Fehr (2004), we capture the existence of reference dependent behavior by the following one-period utility function:

$$(7)\quad v(e_t) = \begin{cases} \lambda(w_t e_t - \tilde{y}) - g(e_t, x_t) & \text{if } w_t e_t \geq \tilde{y} \\ \gamma\lambda(w_t e_t - \tilde{y}) - g(e_t, x_t) & \text{if } w_t e_t < \tilde{y} \end{cases},$$

where $\gamma > 1$ measures the degree of loss aversion, i.e., the increase in the marginal utility of income if the individual is below the income target. Previous evidence (see Kahneman and Tversky 2000) suggests that $\gamma \approx 2$ for many individuals. Loss aversion at this level creates powerful incentives to exert more effort below the income target. Once individuals attain the target $\tilde{y}$, however, the marginal utility of income drops discretely (from $\gamma\lambda$ to λ), causing a substantial reduction in the incentive to supply effort.

The preferences described in (7) imply that workers increase the number of shifts when they are temporarily paid a higher wage: a rise in wages increases the utility of working on a given day. Thus, at higher wages it is more likely that the utility of working $v(e_t)$ exceeds the fixed costs of working. At the same time, however, the increase in wages makes it more likely that the income target is already met or exceeded at relatively low levels of effort. Therefore, compared to the control group, the workers in the treatment group are more likely to face a situation where the marginal utility of income is λ instead of $\gamma\lambda$, i.e., they face lower incentives to work during the shift.[13] As a consequence, members of the treatment group will provide less effort than members of the control group.

The previous discussion shows that reference dependent preferences and a neoclassical model with nonseparable preferences may make similar predictions. In particular, both models are consistent with a reduction in effort per shift during the wage increase. The reduction in effort in the income target model, however, should be related to the degree of loss aversion γ, as explained above. Evidence suggests that there is substantial heterogeneity in the degree of loss aversion between individuals, and that individuals who are loss averse in one type of decisions are also loss averse in other domains of life (see Gaechter, Herrmann, and Johnson 2005). Thus, in principle, the two explanations can be distinguished if one obtains an individual level measure of γ.

[12] See Goette and Huffman (2005) for survey evidence on this point. They present bike messengers with direct survey scenarios to elicit whether messengers care more about making money in the afternoon if they had good luck in the morning than after a bad morning. In their scenarios, good luck means that messengers had the opportunity to make particularly profitable deliveries in the morning. For example, good luck means that a delivery just crosses an additional district boundary; such deliveries command a substantially higher price without much additional effort. About 70 percent of the messengers respond in a fashion consistent with daily income targeting.

[13] If γ is sufficiently high relative to the wage increase, one may obtain the extreme result that the worker provides effort to obtain exactly $\tilde{y}$ before and after the increase. In this case, the worker's effort obviously decreases in response to the wage increase because at higher wages $\tilde{y}$ is obtained at lower effort levels. In general, the larger is γ, the sharper the kink in the objective function and the more likely the worker's optimal effort choice e^* will be at the kink, i.e., the more likely $\gamma\lambda\hat{w}_t > g'(e^*) > \lambda\hat{w}_t$ holds. Note, however, that even if the worker is not a "perfect" income targeter, i.e., even if before or after the wage increase he does not earn exactly $\tilde{y}$, negative effort responses may occur.

TABLE 1—DESCRIPTIVE STATISTICS

		Participating messengers		Difference groups A and B	Nonparticipating messengers, Veloblitz	Messengers, Flash
		Group A	Group B			
Four-week period prior to experiment	Mean revenues	3,500.67	3,269.94	241.67	1461.70	1637.49
		(2,703.25)	(2,330.41)	[563.19]	(1,231.95)	(1,838.61)
	Mean shifts	12.14	10.95	1.20	5.19	6.76
		(8.06)	(7.58)	[1.75]	(4.45)	(6.11)
	N	21	19		21	59
Treatment period 1	Mean revenues	4,131.33	3,005.75	1,125.59	844.21	1,408.23
		(2,669.21)	(2,054.20)	[519.72]	(1,189.53)	(1,664.39)
	Mean shifts	14.00	9.85	4.15	3.14	6.32
		(7.25)	(6.76)	[1.53]	(4.63)	(6.21)
	N	22	20		21	65
Treatment period 2	Mean revenues	2,734.03	3,675.57	−941.53	851.23	921.58
		(2,571.58)	(2,109.19)	[513.2]	(1,150.31)	(1,076.47)
	Mean shifts	8.73	12.55	−3.82	3.29	4.46
		(7.61)	(7.49)	[1.65]	(4.15)	(4.74)
	N	22	20		24	72

Notes: Standard deviations in parentheses, standard error of differences in brackets. Group A received the high commission rate in experimental period 1, group B in experimental period 2.
Source: Own calculations.

III. Results

This section reports the results from our field experiment. Our analysis is based on the four weeks prior to the first experimental period and the two subsequent experimental periods in which first group A and then group B received a wage increase. The data contain the day of each delivery, the messenger's identification number, and the price for each delivery. Thus, we have, in principle, two measures of labor supply: the amount of revenue generated and the number of deliveries completed. Since longer deliveries command a higher price and require more effort, the revenue is our preferred measure of labor supply. Our estimates of the treatment effect, however, are almost identical for either choice of the labor supply measure.

A. *The Impact of the Wage Increase on Total Revenue per Messenger*

The first important question is whether there is a treatment effect on total revenue per messenger during the first and second experimental periods. Tables 1 and 2 present the relevant data. The tables show the revenue data for groups A and B, and the messengers at Flash and Veloblitz who did not participate in the experiment. Table 1 shows the "raw" revenue per messenger—uncontrolled for individual fixed effects. Table 2 controls for individual fixed effects by showing how, on average, the messengers' revenues deviate from their person-specific mean revenues. Thus, a positive number here indicates a positive deviation from the person-specific mean; a negative number indicates a negative deviation.

Tables 1 and 2 show that group A and group B generate very similar revenues per messenger during the four weeks prior to the experiment. If we control for individual fixed effects, we find that the revenues per messenger are almost identical across groups and close to zero. For example, the difference in revenues between group A and group B is only CHF 71.03 if we control for person-specific effects with a standard error of CHF 475.37 (see Table 2). This difference is negligible compared to the average revenue of roughly CHF 3,400 that was generated by a messenger during the preexperimental period. Thus, in the absence of an experimental treatment, the messengers in group A and group B behave in the same way.

During the first experimental period (henceforth, "treatment period 1"), however, in which group A received the higher wage, the total revenue generated by group A is much larger than the revenue of group B, indicating a large treatment effect. On average during this period,

TABLE 2—REVENUES PER FOUR-WEEK PERIOD
(*Average deviations from individual means*)

		Participating messengers			
		Group A	Group B	Nonparticipating messengers, Veloblitz	Messengers, Flash
Four-week period prior to experiment	Mean revenues	−48.88	−119.91	456.72	305.08
		(366.61)	(302.61)	(179.92)	(131.42)
	Difference: group A–group B	71.03			
		(475.37)			
Treatment period 1	Mean revenues	721.98	−277.95	−160.77	102.85
		(192.90)	(240.62)	(173.89)	(105.76)
	Difference: group A–group B	999.93			
		(308.40)			
Treatment period 2	Mean revenues	−675.32	391.87	−258.95	−342.84
		(288.62)	(250.55)	(137.61)	(129.50)
	Difference: group B–group A	1,067.19			
		(382.20)			

Notes: Standard error of the means in parentheses. Same number of observations as in Table 1. Group A received the high commission rate in experimental period 1, group B in experimental period 2.
Source: Own calculations.

messengers in group A generated roughly CHF 4,131 while messengers in group B generated revenues of only CHF 3,006 (see Table 1). This pattern is reversed in the second treatment period, when group B gets the higher wage; group B generates revenues per messenger of CHF 3,676 while group A produces revenues of only CHF 2,734. If we control for individual fixed effects (see Table 2), we can see that the standard errors are relatively small, suggesting that the differences across groups are significant. It is also reassuring that the point estimates of the effects in the two treatment periods are almost identical, pointing to a stable behavioral response to the wage increase.

We perform a statistical test of the effect of the wage increase on revenues in regressions (1)–(3) of Table 3. All regressions are of the form

$$r_{it} = \alpha_i + \delta T_{it} + d_t + e_{it}, \tag{8}$$

where r_{it} measures the revenue generated by messenger i during a four-week period t, α_i is a fixed effect for messenger i, T_{it} is a dummy variable that is equal to 1 if the messenger is on the increased commission rate, d_t is a time dummy estimated for treatment period 1 and for treatment period 2, and e_{it} is the error term.

Regression (1) is based only on the data of groups A and B at Veloblitz. Due to the random assignment of the participating messengers across groups, and due to the fact that both groups served once as a control and once as a treatment group, this regression allows for a clean isolation of the treatment effect. The regression indicates that the treatment effect is highly significant and that the messengers on a high wage generate roughly CHF 1,000 additional revenue compared to the experimental control group.

The two other regressions show that the measured impact of the experimental wage increase on the treated group remains almost the same if we include in the comparison group messengers of Flash and nonparticipants of Veloblitz. Regression (2) compares the treatment group at Veloblitz with *all* other messengers at Veloblitz and finds again a large and significant treatment effect of roughly CHF 1,000. Regression (3) uses observations from all messengers at Veloblitz and the messengers at Flash. The inclusion of the messengers at Flash is suggested by the strong correlation in revenues between Flash and Veloblitz. Regression (3) also includes a dummy for the *whole* nontreated group at Veloblitz, i.e., the messengers in the control group and those who did not participate in the

TABLE 3—MAIN EXPERIMENTAL RESULTS
(*OLS regressions*)

	Dependent variable: Revenues per four-week period			Dependent variable: Shifts per four-week period		
	(1)	(2)	(3)	(4)	(5)	(6)
Observations are restricted to	Messengers participating in experiment	All messengers at Veloblitz	All messengers at Flash and Veloblitz	Messengers participating in experiment	All messengers at Veloblitz	All messengers at Flash and Veloblitz
Treatment dummy	1,033.6***	1,094.5***	1,035.8**	3.99***	4.08***	3.44**
	(326.9)	(297.8)	(444.7)	(1.030)	(0.942)	(1.610)
Dummy for nontreated at Veloblitz			−54.4			−0.772
			(407.4)			(1.520)
Treatment period 1	−211	−370.6	−264.8	−1.28	−1.57	−0.74
	(497.3)	(334.1)	(239.9)	(1.720)	(1.210)	(0.996)
Treatment period 2	−574.7	−656.2	−650.5**	−2.56	−2.63**	−2.19**
	(545.7)	(357.9)	(284.9)	(1.860)	(1.260)	(1.090)
Individual fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R squared	0.74	0.786	0.753	0.694	0.74	0.695
N	124	190	386	124	190	386

Note: Robust standard errors, adjusted for clustering on messengers, are in parentheses.
*** Indicates significance at the 1-percent level.
** Indicates significance at the 5-percent level.
* Indicates significance at the 10-percent level.
Source: Own calculations.

experiment. Therefore, this dummy measures whether the nontreated group at Veloblitz behaved differently relative to the messengers at Flash, and the treatment dummy measures whether the treated group at Veloblitz behaved differently relative to the messengers at Flash. In this regression, the coefficient of the treatment dummy indicates a treatment effect of roughly CHF 1,000. In addition, the dummy for the whole nontreated group at Veloblitz is small and insignificant, indicating that the nontreated group was not affected by the wage increase for the treated group. This result suggests that the wage increase for the treated group did not constrain the opportunities for working for the nontreated group at Veloblitz. The result is also consistent with the permanent existence of unfilled shifts and with survey evidence. The overwhelming majority of the messengers stated that they could work the number of shifts they wanted to work.[14]

In summary, the results above indicate a large and highly significant effect of a temporary wage increase on the total effort of the treated group. In contrast to many previous studies, our results imply a large intertemporal elasticity of substitution. We have seen that the treatment effect is roughly CHF 1,000. The average revenue across group A and group B is CHF 3,568 in treatment period 1; in treatment period 2 it is 3,205. Thus, the intertemporal elasticity of substitution is between (1,000/3,568)/0.25 = 1.12 and (1,000/3,205)/0.25 = 1.25, which is substantially larger compared to what previous studies have found (see, e.g., Oettinger 1999).[15]

[14] It is also noteworthy that we find a negative effect of time on revenues per messenger in all three regressions. The time effect is never significant for the first treatment period, but it is higher for the second treatment period and reaches significance at the 5-percent level in some of the regressions. These time effects suggest that a comparison of the revenues of the same group over time is problematic because revenue is likely to be "polluted" by monthly variations in demand. It is thus not possible to identify the treatment effect by comparing how a group behaved in treatment period 1 relative to treatment period 2.

[15] It is even possible that our measure of the elasticity of labor supply with regard to a temporary wage increase underestimates the true elasticity because we use revenues per messenger as a proxy for labor supply per messenger. If wages w affect effort e and effort affects revenue r, the elasticity of e with respect to w, which we denote by ε_{ew}, is given by $\varepsilon_{rw}/\varepsilon_{re}$, where ε_{rw} is the elasticity of r with respect

TABLE 4—SHIFTS PER FOUR-WEEK PERIOD
(*Average deviations from individual means*)

		Participating messengers		Nonparticipating messengers,	Messengers,
		Group A	Group B	Veloblitz	Flash
Four-week period prior to experiment	Mean shifts	0.22	−0.35	1.57	0.98
		(1.29)	(0.98)	(0.75)	(0.53)
	Difference: group A–group B	0.57			
		(1.62)			
Treatment period 1	Mean shifts	2.53	−1.18	−0.48	0.52
		(0.65)	(0.79)	(0.75)	(0.42)
	Difference: group A–group B	3.71			
		(1.02)			
Treatment period 2	Mean shifts	−2.74	1.52	−0.96	−1.27
		(0.98)	(0.77)	(0.57)	(0.45)
	Difference: group B–group A	4.26			
		(1.24)			

Notes: Standard error of the means in parentheses. Same number of observations as in Table 1. Group A received the high commission rate in experimental period 1, group B in experimental period 2.
Source: Own calculations.

Another common way to calculate this elasticity is to estimate equation (8) in logarithms. Some participants of the experiment, however, did not work at all during the control period and therefore have zero revenues in this four-week period. Hence, taking the logarithm means that these observations have to be removed from the sample. Strictly speaking, then, we would no longer have an experimental comparison.

B. *The Impact of the Wage Increase on Shifts Worked*

After we documented the strong impact of the wage increase on total labor supply, the natural question is whether both the number of shifts and the effort per shift increased. In this section, we examine the impact of the wage increase on the number of shifts worked, while in the next section we take a closer look at effort per shift.

to w (which is observable to us) and ε_{re} is the elasticity of r with respect to e (which is not observable to us). Thus, our measure ε_{rw} implicitly assumes that the elasticity ε_{re} is equal to one. If ε_{re} is less than one, our measure even underestimates the true labor supply elasticity. ε_{re} is less than one if the production function $r = f(e)$ is strictly concave and $f(0) = 0$ holds.

Tables 1 and 4 provide a first indication of how the wage increase affected shifts. Table 1 shows the absolute number of shifts per worker in group A and group B during the four-week period prior to the experiment and the two treatment periods. Table 4 controls for person-specific effects by showing the average deviation of the number of shifts from the person specific means.

Table 1 shows that in the preexperimental period group A worked roughly 12 shifts and group B worked roughly 11 shifts. The standard errors are considerable due to large differences between the workers. If we control for messenger-specific effects (Table 4), we find that the average deviation from person-specific means is very small in both groups and close to zero during the preexperimental period. The deviation from person-specific means is 0.22 in group A (with a standard error of 1.29), and −0.35 in group B (with a standard error of 0.98). Thus, there are almost no differences in shifts across groups before the experiment.

During the first treatment period, however, the messengers in group A, who are paid the high wage, worked almost four shifts more than the messengers in group B (Table 1). Likewise, in the second treatment period the messengers in group B, who now receive the high wage, work four more shifts than the messengers in group B. Moreover, if we control for messenger-specific effects (see Table 4), the standard errors become very small, suggesting that the differences across groups are significant.

A statistical test is presented in regressions (4) through (6) in Table 3. The independent variable in these regressions is s_{it}, the number of shifts that messenger *i* worked during the four-week period *t*. The right-hand side of these regressions is the same as in equation (8), i.e., we included a treatment dummy, individual fixed effects, and time dummies for treatment periods 1 and 2. Regression (4) estimates the impact of the treatment by using only data from group A and group B. It shows a large and highly significant treatment effect; the treated group works on average four shifts more than the control group. Regression (5) uses data from *all* messengers at Veloblitz. The treatment dummy thus compares the treated with the *whole* group of untreated messengers at Veloblitz. This regression basically replicates the results of regression (4). In regression (6), we use data from all messengers at Veloblitz and at Flash. In addition, we include a dummy variable that takes on a value of one if a messenger belongs to the *whole* nontreated group at Veloblitz (which comprises the experimental control group and the messengers who did not participate in the experiment). As in regression (3), this dummy measures whether the experiment had an effect on the whole nontreated group at Veloblitz by comparing this group with Flash messengers. The point estimate on this dummy is small and insignificant, suggesting that the experiment had no effect on the nontreated group at Veloblitz. The treatment dummy in regression (6) compares the treated group with the Flash messengers and again indicates a significant treatment effect of similar size as in the previous regressions.

In summary, regressions (4)–(6) in Table 3 indicate a clear positive treatment effect of the wage increase on shifts. On average, workers supplied about four shifts more if they receive a high commission rate. Since the average number of shifts worked during the two treatment periods is 11.925 and 10.64, respectively, the wage elasticity of shifts is between $(4/11.925)/0.25 = 1.34$ and $(4/10.64)/0.25 = 1.50$. Thus, the shift choices are even more responsive to the wage increase than total revenue per messenger. By definition, the wage elasticity of total revenue is equal to the elasticity of shifts plus the elasticity of the revenue per shift. Therefore, the higher wage elasticity of shifts compared to the elasticity of total revenues is a first indication that the elasticity of effort per shift is negative.

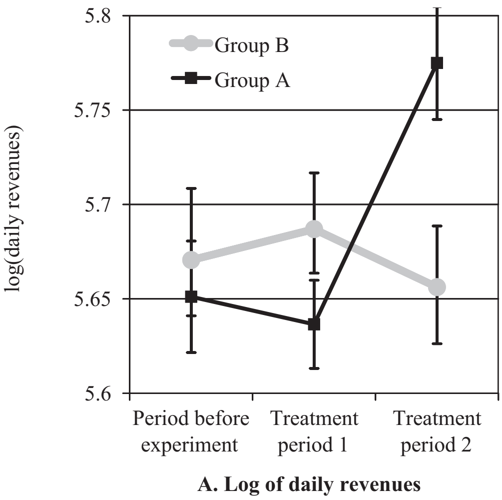
C. *The Impact of the Wage Increase on Effort per Shift*

When examining the revenue per shift, it is useful to restrict attention to behavior during fixed shifts. Recall that the management at Veloblitz did not allow workers to change their fixed shifts after the announcement of the experiment or during the experiment. The increase in the supply of shifts is fully borne by the sign-up shifts. Therefore, our experiment could not induce any kind of selection effect with regard to the fixed shifts and the revenue change during the *fixed* shifts identifies the impact of the treatment on effort per shift.[16]

In Figure 1A, we show the log of revenue per shift in group A and group B during the four weeks prior to the experiment and in the two treatment periods. We control for person effects in Figure 1B by showing the deviation from person-specific means. If we control for person-specific effects, we find that both groups generated almost identical revenues per shift during the four weeks prior to the experiment. During the first treatment period, however, group B, which receives the *lower* wage, generates roughly 5 percent more revenue per shift than group A. Likewise, in the second treatment period, group A, which now receives the *lower* wage, exhibits roughly a 6-percent-higher revenue per shift than group B. Thus, Figure 1 suggests that the wage increase caused a *reduction* in revenue per shift.

The impression conveyed by Figure 1 is further supported by the two regressions presented in Table 5, which are based on observations from group A and group B during fixed shifts. The dependent variable is log revenue of messenger *i* at day *t*. We include a treatment dummy in both regressions that takes on a value of one if mes-

[16] We should, however, mention that the results remain the same when we examine revenue per shift over *all* (fixed and sign-up) shifts.



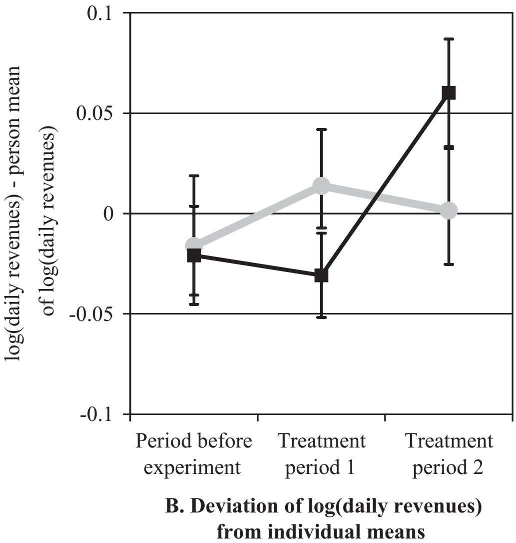


FIGURE 1. LOG OF DAILY REVENUES ON FIXED SHIFTS

Note: Error bars are standard errors of means.

senger i at day t is in the treatment group, and we further control for daily fixed effects and i's tenure. Daily fixed effects are important because of demand variations across days; tenure is important because experienced messengers usually have higher productivity. We do not control for individual fixed effects in regression (1), but for a messenger's gender. This regression shows that the wage increase leads to a reduction in revenue per shift of roughly 6 percent. We control for individual fixed effects in regression (2).

TABLE 5—THE IMPACT OF THE EXPERIMENT ON LOG REVENUES PER DAY
(Dependent variable: log (revenues per shift) during fixed shifts, OLS regressions)

	(1)	(2)
Treatment dummy	−0.0642**	−0.0601**
	(0.030)	(0.030)
Gender (female = 1)	−0.0545	
	(0.052)	
Log(tenure)	0.105***	0.015
	(0.016)	(0.062)
Day fixed effects	Yes	Yes
Individual fixed effects	No	Yes
R-Squared	0.149	0.258
N	1,137	1,137

Note: Robust standard errors, adjusted for clustering on messengers, are in parentheses.
*** Indicates significance at the 1-percent level.
** Indicates significance at the 5-percent level.
* Indicates significance at the 10-percent level.
Source: Own calculations.

The treatment effect in this regression is virtually unchanged and indicates a reduction in revenues of roughly 6 percent.

Thus, the temporary wage increase indeed reduced revenue per shift. The implied wage elasticity of revenue per shift is $-0.06/0.25 = -0.24$, which is consistent with our neoclassical model with preference spillovers across periods and the target income model based on loss aversion. It is also worthwhile to point out that this estimate neatly fills the gap between the elasticity of total revenue and the elasticity of shifts. The intermediate value (between the lower and the upper bound) of the elasticity of total revenue is 1.18. The intermediate value for the elasticity of shifts is 1.42. Thus, according to this difference, the elasticity of effort per shift should be -0.24. Our estimates in Table 5 precisely match this value.

D. *Does Loss Aversion Explain the Negative Impact on Effort per Shift?*

In this section, we provide additional evidence that helps us understand the forces behind the negative impact of the wage increase on effort per shift. Our strategy is to measure individual-level loss aversion and to examine whether these measures have predictive value with regard to individuals' response of effort per shift. In other words, we ask the question whether the loss-averse messengers drive the negative effect of the wage increase on effort per shift or whether the messengers who are not loss averse drive this effect. If mainly the loss-averse messengers show a negative effort response, the loss-aversion model is supported. If the negative effect on effort is not related to individuals' loss aversion, the neoclassical model provides the more plausible explanation.

Loss aversion and reference dependent behavior have implications in a variety of domains. Loss-averse choices have been documented, in particular, in the realm of decision making under uncertainty (Kahneman and Tversky 1979). Therefore, we measured the messengers' loss aversion by observing choices under uncertainty in an experiment that took place eight months after the experimental wage increase. As part of this study, we presented the messengers with the opportunity to participate in the following two lotteries:

Lottery A: Win CHF 8 with probability ½, lose CHF 5 with probability ½. If subjects reject lottery A they receive CHF 0.

Lottery B: This lottery consists of six independent repetitions of lottery A. If subjects reject lottery B they receive CHF 0.

Subjects could participate in both lotteries, or only in one lottery, or they could reject both lotteries.

These lotteries enable us to construct individual measures of loss aversion. In particular, the observed behavior in these lotteries enables us to classify subjects with regard to their degree of loss aversion γ. If subjects' preferences are given by (7), subjects who reject lottery A have a higher level of γ than subjects who accept lottery A, and subjects who reject lottery A *and* B have a higher level of γ than subjects who reject only lottery A. In addition, if subjects' loss aversion is consistent across the two lotteries, then any individual who rejects lottery B should also reject lottery A because a rejection of lottery B implies a higher level of loss aversion than a rejection of only lottery A. We derive these implications of (7) explicitly in Appendix A.

Among the 42 messengers who belong to either group A or group B, 19 messengers rejected both lotteries, 8 messengers rejected only lottery A, 1 messenger rejected only lottery B, and 14 messengers accepted both lotteries. Thus, with the exception of the one messenger who rejects only lottery B, all messengers who rejected lottery B also rejected lottery A. These results can be neatly captured by a simple loss-averse utility function that obeys equation (7).[17]

In principle, one might think that the rejection of A and/or B is also compatible with risk aversion arising from diminishing marginal utility of lifetime income. Matthew Rabin's calibration theorem (Rabin 2000) rules out this interpretation, however. Rabin showed that a theory of risk-averse behavior based on the assumption of diminishing marginal utility of *lifetime* income implies that people essentially *must be* risk neutral for low-stake gambles like our lotteries. Intuitively, this follows from the fact that risk-averse behavior for low-stake gambles implies ridiculously high levels of risk aversion for slightly higher, but still moderate, stake levels. Yet, such unreasonably high levels of risk aversion can be safely ruled out. For example, we show in Appendix B that if one assumes that the rejection of lottery A is driven by diminishing marginal utility of lifetime income, then the subject will also reject a lottery where one can lose \$32 with probability ½ and win *any* positive prize with probability ½. Thus, there is no finite prize that induces this subject to accept a 50-percent chance of losing \$32. Similar results are implied by a rejection of lottery B.

In Figure 2, we illustrate the behavior of messengers with and without loss-averse preferences. The figure controls for person-specific effects by comparing individual log revenues to the mean of the individual's log revenues. We show that the messengers who did not display loss-averse preferences do not change their effort per shift across the treatment and the control period. The messengers who displayed loss aversion in the lottery choices, however, exhibit a lower effort per shift in the treatment period

[17] These results are qualitatively similar to the results obtained in a many other studies (e.g., Daniel Read, George Loewenstein, and Matthew Rabin 1999; Robin Cubbit, Chris Starmer, and Robert Sugden 1998; Robin Hogarth and Hillel Einhorn 1992; Gideon Keren and Willem Wagenaar 1987).

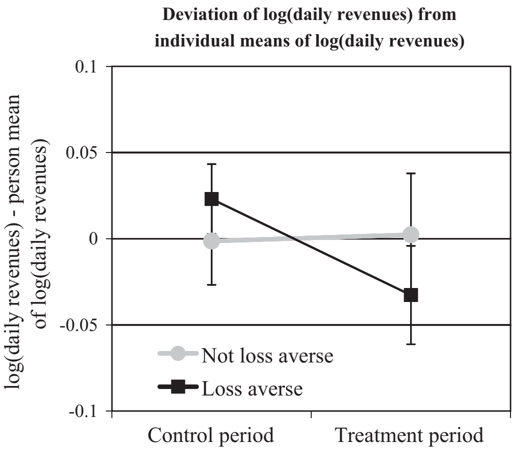


FIGURE 2. THE BEHAVIOR OF LOSS-AVERSE AND NOT-LOSS-AVERSE SUBJECTS DURING CONTROL AND TREATMENT PERIOD IN FIXED SHIFTS

Note: Error bars are standard errors of means.

TABLE 6—DOES LOSS AVERSION EXPLAIN THE TREATMENT EFFECT?
(*Dependent variable: log (revenues per shift) during fixed shifts, OLS regressions*)

	(1)	(2)
Treatment effect × not loss averse	−0.0273	−0.027
	(0.033)	(0.032)
Treatment effect × rejects lottery A	−0.105**	
	(0.046)	
Treatment effect × rejects one lottery		−0.0853*
		(0.062)
Treatment effect × rejects both lotteries		−0.12**
		(0.053)
Log(tenure)	0.00152	0.0074
	(0.061)	(0.060)
Day fixed effects	Yes	Yes
Individual fixed effects	Yes	Yes
R-Squared	0.243	0.26
N	1137	1137

Note: Robust standard errors, adjusted for clustering on messengers, are in parentheses.
*** Indicates significance at the 1-percent level.
** Indicates significance at the 5-percent level.
* Indicates significance at the 10-percent level.
Source: Own calculations.

compared to the control period. This pattern suggests that the negative effect of wages on effort per shift may be driven solely by the loss-averse messengers.

To examine this possibility in more depth, we ran the regressions in Table 6. In these regressions, log daily revenue of messenger *i* at day *t* is again the dependent variable and we control for messenger fixed effects in all regressions, as loss-averse messengers may differ in more than one dimension from other messengers. In the first regression, we split the treatment group according to behavior in lottery A. If a messenger rejects lottery A, the messenger is more loss averse than if lottery A is accepted. In regression (1), we estimate the treatment effect separately for loss-averse messengers (who rejected lottery A) and messengers who did not display loss aversion (who accepted lottery A). The results show that loss-averse messengers generated roughly 10-percent lower revenue per shift when they received the high wage. In contrast, the treatment effect is much lower and insignificant for the messengers without loss aversion.

Regression (2) of Table 6 provides a robustness check for this result because we use a finer scale of messengers' loss aversion which yields treatment effects for three separate groups: messengers accepting both lotteries (labeled "not loss averse"), messengers rejecting one of the two lotteries, and messengers rejecting both lotteries. The theory predicts that the strongest treatment effect should occur for the group that rejects both lotteries, followed by the group that rejects only one lottery. We do find evidence of this, although the differences between those who reject both and those who reject only one lottery are small. Regression (2) also shows that the wage increase triggers no significantly negative impact on messengers who exhibit no loss aversion in the lotteries, while the other two groups exhibit clear reductions in revenues during the wage increase. These results suggest that the negative impact of the wage increase on revenue per shift is associated with the messengers' degree of loss aversion, lending support to the target income model discussed in Section IIC.

V. Summary

This paper reports the results of a randomized field experiment examining how workers, who can freely choose their working time, and their effort during working time, respond to a fully anticipated temporary wage increase. We find a strong positive impact of the wage increase on total labor supply during the two four-week periods in which the experiment took place. The associated intertemporal elasticity of substitution is between 1.12 and 1.25. The large increase in total labor supply is exclusively driven by the increase in the number of shifts worked. On average, messengers increase their working time during the four weeks in which they receive a higher wage by four shifts (20 hours), which implies a wage elasticity of shifts between 1.34 and 1.50. This is a considerably larger elasticity than what has previously been found on the basis of daily labor supply data (Camerer et al. 1997; Chou 2002; Oettinger 1999). We also find that the wage increase causes a decrease in revenue (effort) per shift by roughly 6 percent. The increase in the number of shifts, however, dominates the negative impact on effort per shift by a large margin such that overall labor supply strongly increases.

The standard neoclassical model with separable intertemporal utility is not consistent with the evidence because this model predicts that both the number of shifts and the effort per shift increase in response to the wage increase. We show, however, that a neoclassical model with preference spillovers across periods as well as a target income model with loss-averse preferences are consistent with the observed decrease in effort per shift. In order to discriminate between these two models, we measured the messengers' loss aversion at the individual level in the domain of choices under uncertainty. We use these measures to examine whether the negative impact of the wage increase on effort per shift is mediated by the degree to which messengers' are loss averse. We find that the degree of loss aversion is indeed related to the response of effort per shift. Higher degrees of loss aversion are associated with a stronger negative impact of the wage increase on effort per shift, and workers who do not display loss aversion in choices under uncertainty also do not show a significant effort reduction. Thus, it seems that loss aversion drives the negative effect of wages on effort.

We believe that these results contribute to a deeper understanding of the behavioral foundations of labor supply. Our results do not rule out a role for "neoclassical" preferences in labor supply decisions. One-third of the workers in our sample did not exhibit loss aversion, and the large intertemporal substitution effects on overall labor supply and the supply of shifts document the power of behavioral forces that have always been emphasized in the standard life-cycle model. Our results also contrast sharply with the small and insignificant substitution effects that have been found in many previous studies. Therefore, the small effects in these studies may reflect the constraints workers face in their labor supply decisions and—in view of our results—may be less likely due to workers' unwillingness to substitute labor *hours* over time. Our results on the behavioral sources of the negative wage elasticity of effort per shift also suggest, however, that disregarding reference dependent preferences in *effort* decisions is not wise.

APPENDIX A

In this appendix, we derive the conditions under which a loss-averse individual, whose preferences obey (7) in the text, will reject lotteries A and B. For the purpose of lottery choices, the disutility of effort does not matter so that we can simplify preferences to

$$v(x-r)=\begin{cases}\lambda(x-r) & \text{if } x\geq r\\ \gamma\lambda(x-r) & \text{if } x<r\end{cases},$$

where x is the lottery outcome and r is the reference point. We take the reference point to be the status quo. The individual will reject gamble A if

$$0.5v(-5)+0.5v(8)\leq v(0),$$

which simplifies to

$$0.5(-5\gamma\lambda)+0.5(8)\lambda\leq 0.$$

This condition is satisfied if

$$\gamma\geq\frac{8}{5}.$$

The individual will reject gamble B if

$$\frac{1}{64}v(-30)+\frac{6}{64}v(-17)+\frac{15}{64}v(-4)+\frac{20}{64}v(9)$$
$$+\frac{15}{64}v(22)+\frac{6}{64}v(35)+\frac{1}{64}v(48)\leq v(0).$$

Plugging in our functional form and simplifying, we find that the individual will reject the gamble if

$$\gamma \geq \frac{793}{192}.$$

As claimed in the text, the degree of loss aversion required to reject gamble B is greater than the degree of loss aversion needed to reject A.

Appendix B

In this appendix, we prove the following: If an individual has a concave utility function $u(\ )$ and rejects a coin flip, where she can either win CHF 8 or lose CHF 5, for all wealth levels (m, ∞), then she will reject *any* coin flip in which she could lose CHF 32 no matter how large the positive prize that is associated with the coin flip.

PROOF:

We proceed in four steps:

(i) We adopt the convention that, if indifferent, the individual rejects the coin flip. Rejecting the coin flip implies

$$0.5u(m+8) + 0.5u(m-5) \leq u(m)$$

$$\Leftrightarrow u(m+8) - u(m) \leq u(m) - u(m-5).$$

It follows from concavity that $8[u(m+8) - u(m+7)] \leq u(m+8) - u(m)$ and $u(m) - u(m-5) \leq 5[u(m-4) - u(m-5)]$. Define $MU(x) = u(x) - u(x-1)$ as the marginal utility of the xth dollar. Putting the last three inequalities together, we obtain

$$MU(m+8) \leq \frac{5}{8} MU(m-5)$$

and, because of the premise, it is true that

$$MU(x+12) \leq \frac{5}{8} MU(x) \text{ for all } x > m-4.$$

(ii) We now derive an upper bound on $u(\infty)$. The concavity of $u(\ )$ implies

$$u(m+12) \leq u(m) + 12MU(m).$$

Using the same logic,

$$u(m+24) \leq u(m) + 12MU(m) + 12MU(m+12)$$

$$\leq u(m) + 12MU(m)\left[1 + \frac{5}{8}\right]$$

$$u(m+36) \leq u(m) + 12MU(m)\left[1 + \frac{5}{8} + \left(\frac{5}{8}\right)^2\right]$$

and so on. Thus, we can develop a geometric series starting from m. Taking the limit, we obtain

$$u(\infty) \leq u(m) + 12MU\frac{8}{3} = u(m) + 32MU(m).$$

(iii) Concavity implies $u(m-32) \leq u(m) - 32MU(m)$.

(iv) Using the results from steps (ii) and (iii), we get an upper bound on the value of a coin flip where the individual would either lose CHF 32 or win an infinite amount:

$$0.5u(m-32) + 0.5u(\infty) \leq u(m).$$

This implies that the individual would reject the gamble.

REFERENCES

Altonji, Joseph G. 1986. "Intertemporal Substitution in Labor Supply: Evidence from Micro Data." *Journal of Political Economy*, 94(3): S176–215.

Blundell, Richard. 1994. "Evaluating Structural Microeconometric Models of Labor Supply." In *Advances in Econometrics: Sixth World Congress of the Econometric Society Vol. II*, ed. Christopher A. Sims, 3–48. Cambridge: Cambridge University Press.

Blundell, Richard, and Thomas Macurdy. 1999. "Labor Supply: A Review of Alternative Approaches." In *Handbook of Labor Economics Volume 3A*, ed. Orley Ashenfelter and

DavidCard, 1559–1695. Amsterdam, New York, and Oxford: Elsevier Science.

Browning, Martin, Angus Deaton, and Margaret Irish. 1985. "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle." *Econometrica*, 53(3): 503–43.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor Supply of New York City Cab Drivers: One Day at a Time." *Quarterly Journal of Economics*, 112(2): 407–441.

Card, David. 1994. "Intertemporal Labor Supply: An Assessment." In *Advances in Econometrics*: *Sixth World Congress of the Econometric Society Vol. II*, ed. Christopher A. Sims, 49–78. Cambridge: Cambridge University Press.

Chou, Yuan K. 2002. "Testing Alternative Models of Labour Supply: Evidence from Taxi Drivers in Singapore." *Singapore Economic Review*, 47(1): 17–47.

Cubitt, Robin P., Chris Starmer, and Robert Sugden. 1998. "On the Validity of the Random Lottery Incentive System." *Experimental Economics*, 1(2): 115–31.

Dickens, William T., and Shelly J. Lundberg. 1993. "Hours Restrictions and Labor Supply." *International Economic Review*, 34(1): 169–92.

Farber, Henry S. 2004. "Reference-Dependent Preferences and Labor Supply: The Case of New York City Cabdrivers." Princeton University Working Paper 497.

Farber, Henry S. 2005. "Is Tomorrow Another Day? The Labor Supply of New York City Cab Drivers." *Journal of Political Economy*, 113(1): 46–82.

Gaechter, Simon, Andreas Hermann, and Eric Johnson. 2005. "Individual-Level Loss Aversion in Risky and Riskless Choice." Unpublished.

Goette, Lorenz, David Huffman, and Ernst Fehr. 2004. "Loss Aversion and Labor Supply." *Journal of the European Economic Association*, 2(2–3): 216–28.

Goette, Lorenz, and David Huffman. 2005. "Affect as a Source of Motivation in the Workplace: A New Model of Labor Supply, and New Field Evidence on Income Targeting and the Goal Gradient." Institute for the Study of Labor Discussion Paper 1890.

Heath, Chip, Richard P. Larrick, and George Wu. 1999. "Goals as Reference Points." *Cognitive Psychology*. 38(1): 79–109.

Hogarth, Robin M., and Hillel J. Einhorn. 1990. "Venture Theory: A Model of Decision Weights." *Management Science*. 36(7): 780–803.

Kahn, Shulamit, and Kevin Lang. 1991. "The Effect of Hours Constraints on Labor Supply Estimates." *Review of Economics and Statistics*, 73(4): 605–11.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47(2): 263–91.

Kahneman, Daniel, and Amos Tversky. 2000. *Choices, Values, and Frames*. Cambridge, New York, and Melbourne: Cambridge University Press.

Keren, Gideon, and Willem Wagenaar. 1987. "Violation of Utility Theory in Unique and Repeated Gambles." *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 13(3): 387–391.

Mankiw, N. Gregory, Julio J. Rotemberg, and Lawrence H. Summers. 1985. "Intertemporal Substitution in Macroeconomics." *Quarterly Journal of Economics*, 100(1): 225–51.

Oettinger, Gerald S. 1999. "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors." *Journal of Political Economy*, 107(2): 360–92.

Pencavel, John H. 1986. "Labor Supply of Men: A Survey." In *Handbook of Labor Economics Volume 1*, ed. Orley Ashenfelter and Richard Layard, 3–102. Amsterdam, New York, and Oxford: Elsevier Science.

Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica*, 68(5): 1281–92.

Read, Daniel, George Loewenstein, and Matthew Rabin. 1999. "Choice Bracketing." *Journal of Risk and Uncertainty*, 19(1-3): 171–97.

Romer, David. 1996. *Advanced Macroeconomics*. New York: McGraw-Hill.

This article has been cited by:

1. Yingchao Zhang, Oliver Fabel, Christian Thomann. 2015. Pay inequity effects on back-office employees' job performances: the case of a large insurance firm. *Central European Journal of Operations Research* **23**, 421-439. [CrossRef]
2. Oriana Bandiera, Valentino Larcinese, Imran Rasul. 2015. Blissful ignorance? A natural experiment on the effect of feedback on students' performance. *Labour Economics* **34**, 13-25. [CrossRef]
3. Hang Ye, Shu Chen, Daqiang Huang, Siqi Wang, Yongmin Jia, Jun Luo. 2015. Transcranial direct current stimulation over prefrontal cortex diminishes degree of risk aversion. *Neuroscience Letters* **598**, 18-22. [CrossRef]
4. Heiko Karle, Georg Kirchsteiger, Martin Peitz. 2015. Loss Aversion and Consumption Choice: Theory and Experimental Evidence. *American Economic Journal: Microeconomics* **7**:2, 101-120. [Abstract] [View PDF article] [PDF with links]
5. Stefan Moser, Oliver Mußhoff. 2015. Ex-ante Evaluation of Policy Measures: Effects of Reward and Punishment for Fertiliser Reduction in Palm Oil Production. *Journal of Agricultural Economics* n/a-n/a. [CrossRef]
6. Lukas M. Wenner. 2015. Expected prices as reference points—Theory and experiments. *European Economic Review* **75**, 60-79. [CrossRef]
7. Florian Zimmermann. 2015. Clumped or Piecewise? Evidence on Preferences for Information. *Management Science* **61**, 740-753. [CrossRef]
8. Andreas Ortmann. 2015. Book Review. *Journal of Economic Psychology* . [CrossRef]
9. Björn Bartling, Leif Brandes, Daniel Schunk. 2015. Expectations as Reference Points: Field Evidence from Professional Soccer. *Management Science* 150129074431005. [CrossRef]
10. Matthew Harding, Alice Hsiaw. 2014. Goal setting and energy conservation. *Journal of Economic Behavior & Organization* **107**, 209-227. [CrossRef]
11. SUNG-HA HWANG, SAMUEL BOWLES. 2014. Optimal Incentives with State-Dependent Preferences. *Journal of Public Economic Theory* **16**, 681-705. [CrossRef]
12. Thomas Dohmen. 2014. Behavioral labor economics: Advances and future directions. *Labour Economics* **30**, 71-85. [CrossRef]
13. David Eil, Jaimie W. Lien. 2014. Staying ahead and getting even: Risk attitudes of experienced poker players. *Games and Economic Behavior* **87**, 50-69. [CrossRef]
14. T. Buser, M. Niederle, H. Oosterbeek. 2014. Gender, Competitiveness, and Career Choices. *The Quarterly Journal of Economics* **129**, 1409-1447. [CrossRef]
15. Alain Cohn, Ernst Fehr, Benedikt Herrmann, Frédéric Schneider. 2014. SOCIAL COMPARISON AND EFFORT PROVISION: EVIDENCE FROM A FIELD EXPERIMENT. *Journal of the European Economic Association* **12**:10.1111/jeea.2014.12.issue-4, 877-898. [CrossRef]
16. Arndt Werner, Johanna Gast, Sascha Kraus. 2014. The effect of working time preferences and fair wage perceptions on entrepreneurial intentions among employees. *Small Business Economics* **43**, 137-160. [CrossRef]
17. Heiko Karle, Martin Peitz. 2014. Competition under consumer loss aversion. *The RAND Journal of Economics* **45**:10.1111/rand.2014.45.issue-1, 1-31. [CrossRef]
18. Alain Cohn, Ernst Fehr, Lorenz Goette. 2014. Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment. *Management Science* 141223041315002. [CrossRef]

19. Ola Andersson, Håkan J. Holm, Jean-Robert Tyran, Erik Wengström. 2014. Deciding for Others Reduces Loss Aversion. *Management Science* 141223041315002. [CrossRef]
20. Kirk Doran. 2014. Are long-term wage elasticities of labor supply more negative than short-term ones?. *Economics Letters* **122**, 208-210. [CrossRef]
21. Craig R. Fox, Russell A. PoldrackProspect Theory and the Brain 533-567. [CrossRef]
22. 2014. The Intrinsic Value of Decision Rights. *Econometrica* **82**, 2005-2039. [CrossRef]
23. Tom Chang, Tal Gross. 2013. How Many Pears Would a Pear Packer Pack if a Pear Packer CouldPack Pears at Quasi-Exogenously Varying Piece Rates?. *Journal of Economic Behavior & Organization* . [CrossRef]
24. Elaine M. Liu. 2013. Time to Change What to Sow: Risk Preferences and Technology Adoption Decisions of Cotton Farmers in China. *Review of Economics and Statistics* **95**, 1386-1403. [CrossRef]
25. Quang Nguyen, Pingsun Leung. 2013. Revenue targeting in fisheries. *Environment and Development Economics* **18**, 559-575. [CrossRef]
26. Alisa G. Brink, Frederick W. Rankin. 2013. The Effects of Risk Preference and Loss Aversion on Individual Behavior under Bonus, Penalty, and Combined Contract Frames. *Behavioral Research in Accounting* **25**, 145-170. [CrossRef]
27. E. Dal Bo, F. Finan, M. A. Rossi. 2013. Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service. *The Quarterly Journal of Economics* **128**, 1169-1218. [CrossRef]
28. Elaine M. Liu, JiKun Huang. 2013. Risk preferences and pesticide use by cotton farmers in China. *Journal of Development Economics* **103**, 202-215. [CrossRef]
29. Håkan Eggert, Viktoria Kahui. 2013. Reference-dependent behaviour of paua (abalone) divers in New Zealand. *Applied Economics* **45**, 1571-1582. [CrossRef]
30. Stefan T. Trautmann, Razvan Vlahu. 2013. Strategic loan defaults and coordination: An experimental analysis. *Journal of Banking & Finance* **37**, 747-760. [CrossRef]
31. Alice Hsiaw. 2013. Goal-setting and self-control. *Journal of Economic Theory* **148**, 601-626. [CrossRef]
32. Ian Larkin, Lamar Pierce, Francesca Gino. 2012. The psychological costs of pay-for-performance: Implications for the strategic compensation of employees. *Strategic Management Journal* **33**:10.1002/smj.v33.10, 1194-1214. [CrossRef]
33. David A. Spencer. 2012. Barbarians at the gate: a critical appraisal of the influence of economics on the field and practice of HRM. *Human Resource Management Journal* no-no. [CrossRef]
34. Tomer Blumkin, Bradley J. Ruffle, Yosef Ganun. 2012. Are income and consumption taxes ever really equivalent? Evidence from a real-effort experiment with real goods. *European Economic Review* **56**, 1200-1219. [CrossRef]
35. Esther Duflo,, Rema Hanna,, Stephen P. Ryan. 2012. Incentives Work: Getting Teachers to Come to School. *American Economic Review* **102**:4, 1241-1278. [Abstract] [View PDF article] [PDF with links]
36. Samuel Bowles,, Sandra Polanía-Reyes. 2012. Economic Incentives and Social Preferences: Substitutes or Complements?. *Journal of Economic Literature* **50**:2, 368-425. [Abstract] [View PDF article] [PDF with links]
37. Paul Dolan, Antony Elliott, Robert Metcalfe, Ivo Vlaev. 2012. Influencing Financial Behavior: From Changing Minds to Changing Contexts. *Journal of Behavioral Finance* **13**, 126-142. [CrossRef]
38. David Gill,, Victoria Prowse. 2012. A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review* **102**:1, 469-503. [Abstract] [View PDF article] [PDF with links]
39. P. Edwards. 2011. Experimental economics and workplace behaviour: bridges over troubled methodological waters?. *Socio-Economic Review* . [CrossRef]

40. Alex Bryson, Erling Barth, Harald Dale-Olsen. 2011. Do Higher Wages Come at a Price?. *Journal of Economic Psychology* . [CrossRef]

41. P. Dolan, M. Hallsworth, D. Halpern, D. King, R. Metcalfe, I. Vlaev. 2011. Influencing behaviour: the mindspace way. *Journal of Economic Psychology* . [CrossRef]

42. Vincent P. Crawford,, Juanjuan Meng. 2011. New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income. *American Economic Review* **101**:5, 1912-1932. [Abstract] [View PDF article] [PDF with links]

43. David Card,, Stefano DellaVigna,, Ulrike Malmendier. 2011. The Role of Theory in Field Experiments. *Journal of Economic Perspectives* **25**:3, 39-62. [Abstract] [View PDF article] [PDF with links]

44. RICHARD ROGERSON. 2011. Individual and Aggregate Labor Supply with Coordinated Working Times. *Journal of Money, Credit and Banking* **43**:10.1111/jmcb.2011.43.issue-s1, 7-37. [CrossRef]

45. Peter A. Bibby, Eamonn Ferguson. 2011. The ability to process emotional information predicts loss aversion. *Personality and Individual Differences* **51**, 263-266. [CrossRef]

46. Stefan T. Trautmann, Ferdinand M. Vieider, Peter P. Wakker. 2011. Preference Reversals for Ambiguity Aversion. *Management Science* **57**, 1320-1333. [CrossRef]

47. JUSTIN ESAREY, TIMOTHY C. SALMON, CHARLES BARRILLEAUX. 2011. WHAT MOTIVATES POLITICAL PREFERENCES? SELF-INTEREST, IDEOLOGY, AND FAIRNESS IN A LABORATORY DEMOCRACY. *Economic Inquiry* no-no. [CrossRef]

48. Edith Law, Luis von Ahn. 2011. Human Computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning* **5**, 1-121. [CrossRef]

49. Rick K. Wilson. 2011. The Contribution of Behavioral Economics to Political Science. *Annual Review of Political Science* **14**, 201-223. [CrossRef]

50. R. Chetty, J. N. Friedman, T. Olsen, L. Pistaferri. 2011. Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *The Quarterly Journal of Economics* **126**, 749-804. [CrossRef]

51. Johannes Abeler,, Armin Falk,, Lorenz Goette,, David Huffman. 2011. Reference Points and Effort Provision. *American Economic Review* **101**:2, 470-492. [Abstract] [View PDF article] [PDF with links]

52. Hans-Martin von Gaudecker,, Arthur van Soest,, Erik WengstrÖm. 2011. Heterogeneity in Risky Choice Behavior in a Broad Population. *American Economic Review* **101**:2, 664-694. [Abstract] [View PDF article] [PDF with links]

53. Corina Paraschiv, Regis Chenavaz. 2011. Sellers' and Buyers' Reference Point Dynamics in the Housing Market. *Housing Studies* **26**, 329-352. [CrossRef]

54. Peter Chinloy, Daniel T. Winkler. 2011. Contracts, Labor Supply and Income Targeting. *Journal of Labor Research* . [CrossRef]

55. Alexander K. Koch, Julia Nafziger. 2011. Self-regulation through Goal Setting*. *Scandinavian Journal of Economics* **113**:10.1111/sjoe.2011.113.issue-1, 212-227. [CrossRef]

56. Devin G. Pope,, Maurice E. Schweitzer. 2011. Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes. *American Economic Review* **101**:1, 129-157. [Abstract] [View PDF article] [PDF with links]

57. Yves Zenou. 2011. RURAL-URBAN MIGRATION AND UNEMPLOYMENT: THEORY AND POLICY IMPLICATIONS*. *Journal of Regional Science* **51**:10.1111/jors.2011.51.issue-1, 65-82. [CrossRef]

58. Gary Charness, Peter KuhnLab Labor: What Can Labor Economists Learn from the Lab? 229-330. [CrossRef]

59. James B. Rebitzer, Lowell J. TaylorExtrinsic Rewards and Intrinsic Motives: Standard and Behavioral Approaches to Agency and Labor Markets 701-772. [CrossRef]

60. Nava Ashraf,, James Berry,, Jesse M. Shapiro. 2010. Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia. *American Economic Review* **100**:5, 2383-2413. [Abstract] [View PDF article] [PDF with links]

61. Randolph Sloof, C. Mirjam van Praag. 2010. The effect of noise in a performance measure on work motivation: A real effort laboratory experiment. *Labour Economics* **17**, 751-765. [CrossRef]

62. Heike Hennig-Schmidt, Abdolkarim Sadrieh, Bettina Rockenbach. 2010. In Search of Workers' Real Effort Reciprocity-a Field and a Laboratory Experiment. *Journal of the European Economic Association* **8**:10.1111/jeea.2010.8.issue-4, 817-837. [CrossRef]

63. Joshua D. Angrist,, Jörn-Steffen Pischke,. 2010. The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives* **24**:2, 3-30. [Abstract] [View PDF article] [PDF with links]

64. Erica Mina Okada. 2010. Uncertainty, Risk Aversion, and WTA vs. WTP. *Marketing Science* **29**, 75-84. [CrossRef]

65. Stefano DellaVigna. 2009. Psychology and Economics: Evidence from the Field. *Journal of Economic Literature* **47**:2, 315-372. [Abstract] [View PDF article] [PDF with links]

66. Craig R. Fox, Russell A. PoldrackProspect Theory and the Brain 145-173. [CrossRef]

67. Gelkha Buitrago, Werner Gueth, Maria Vittoria Levati. 2009. On the relation between impulses to help and causes of neediness: An experimental study. *The Journal of Socio-Economics* **38**, 80-88. [CrossRef]

68. Oded Netzer, Olivier Toubia, Eric T. Bradlow, Ely Dahan, Theodoros Evgeniou, Fred M. Feinberg, Eleanor M. Feit, Sam K. Hui, Joseph Johnson, John C. Liechty, James B. Orlin, Vithala R. Rao. 2008. Beyond conjoint analysis: Advances in preference measurement. *Marketing Letters* **19**, 337-354. [CrossRef]

69. Rajeev Dehejia, Bjorn Jorgensen, Raphael Thomadsen. 2008. Optimal Minimum Wage in the Classic Labor Supply-and-Demand Paradigm. *Journal of Poverty* **12**, 481-495. [CrossRef]

70. Henry S. Farber. 2008. Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers. *American Economic Review* **98**:3, 1069-1082. [Abstract] [View PDF article] [PDF with links]

71. Michael Conlin, Ted O'Donoghue, Timothy J. Vogelsang. 2007. Projection Bias in Catalog Orders. *American Economic Review* **97**:4, 1217-1249. [Abstract] [View PDF article] [PDF with links]

72. Glenn W. Harrison, E. Elisabet RutströmRisk Aversion in the Laboratory 41-196. [CrossRef]