Labour Economics 88 (2024) 102533

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Working more for more and working more for less: Labor supply in the gain and loss domains

C. Bram Cadsby [a,1], Fei Song [b], Nick Zubanov [c,*]

[a] University of Guelph, Canada
[b] Toronto Metropolitan University, Canada
[c] University of Konstanz, Germany

ARTICLE INFO

JEL:
D91
J22
J31
M52

Keywords:
Labor supply
Framing
Reference points
Loss aversion
Experiment

ABSTRACT

We examine labor supply responses to piece rate changes relative to the reference piece rate (RR). In experimental conditions without RR, labor supply increases monotonically with the actual piece rate. In conditions with RR, labor supply increases both when the piece rate rises and falls relative to RR. This non-monotonicity in labor supply responses to piece rate changes around RR is consistent with the effects of framing a given level of income as gain or loss relative to the target level induced by RR: loss aversion makes subjects work more at a given piece rate when the implied income is in the loss rather than gain domain. However, the framing effects disappear when the piece rate could both rise or fall relative to RR.

1. Introduction

Reference dependence, the key concept in Prospect Theory (Kahneman and Tversky, 1979), has been applied to study decisions in numerous domains including household savings and finance (Camerer, 2004), consumer behavior (Ericson and Fuster, 2011; Hardie et al., 1993), environmental protection and food safety (Milkman et al., 2012), fairness and justice judgments (Ganegoda and Folger, 2015), status concerns (Polman, 2012), and, directly relevant to this study, labor supply (e.g., Hossain and List, 2012; Abeler et al., 2011; Andersen et al., 2014). We apply this concept to experimentally examine the response of labor supply to a given piece rate change in the presence of the varying reference piece rate (RR). The prospect theory implies that, as long as the reference, or "target", income is related to RR that generates it, changes in RR should create framing effects akin to those caused by labor income being in the gain versus loss domain relative to the target. Hence our research question: does labor supply respond differently to a given piece rate change when it is framed as a loss versus gain relative to RR?

Labor supply reactions to changes in the price of labor, whether paid per unit of time or, as in our study, per unit of output,[2] are important to many research fields ranging from personnel management to public finance. Within the neoclassical model of labor supply, when piece rate increases, the extra income earned through working creates an incentive to consume more leisure and hence work less (the income effect), but at the same time, the now higher opportunity cost of leisure induces substitution away from leisure towards more work (the substitution effect). Focusing on short-term changes in piece rates – a setting frequently occurring in empirical labor supply studies, and replicated in ours, – the neoclassical model would predict labor supply to monotonically increase with piece rate (e.g., Fehr and Goette, 2007, Section II), owing to intertemporal substitution between consumption (funded by labor income) and leisure. Indeed, several studies find that people work more for higher piece rates, and in diverse work contexts: construction of the Trans-Alaskan pipeline (Carrington, 1996), vending at sports stadiums (Oettinger, 1999), planting trees (Shearer, 2004), installing windscreens (Lazear, 2000), or harvesting lobsters (Stafford, 2015).

Importantly, the neoclassical prediction of labor supply

* Corresponding author.
E-mail address: nick.zubanov@uni-konstanz.de (N. Zubanov).

[1] C. Bram Cadsby died in February 2022.

[2] While piece rate and hourly wage rate may reflect different dimensions of labor supply, abstracting from agency issues in effort provision, as we do in this study, makes these two concepts nearly identical.

https://doi.org/10.1016/j.labeco.2024.102533
Received 30 December 2021; Received in revised form 22 January 2024; Accepted 29 February 2024
Available online 7 March 2024

monotonically increasing with piece rate in the short run also implies that people should work *less* for lower piece rates. However, empirical support for this proposition is not unanimous. A seminal study by Camerer et al. (1997) finds that New York City taxi drivers drove more hours on days when their hourly earnings were lower – an unexpectedly strong income effect dominating intertemporal substitution between labor and leisure given that hourly earnings fluctuated daily. The authors applied prospect theory to explain their results through income targeting: earning an income below a certain reference level, or "target", is framed as a loss, which loss-averse drivers try to avoid by working more hours. The income target being reached less quickly on less lucrative days explains the puzzle of working more for less.

Though Camerer et al. (1997) study has been highly influential, follow-up field studies produced mixed results (see Stafford, 2018, for a survey). Some found evidence consistent with framing effects (e.g., Chou, 2002 for taxi drivers in Singapore; Chang and Gross, 2014 for fruit pickers in California; Della Vigna et al., 2017 for job search effort of the unemployed in Hungary); others found only weak or no evidence (Farber, 2005, 2008, 2015, using later and more complete NYC taxi driver records); while yet other studies reconciled those conflicting findings by allowing for targets based on work hours as well as income (Crawford and Meng, 2011, for NYC taxi drivers).

A related experimental literature, to which our work belongs, studies labor supply under compensation schemes in which the income under a given piece (or hourly) rate is framed as a gain or loss relative to a pre-existing or experimentally induced income target. Such framing matters: for example, Brownback and Sadoff (2020) find that people prefer gain-framed incentives to loss-framed incentives of the same expected monetary value. However, whether framing influences labor supply study as prospect theory would predict is debatable. Some experiments find labor supply to respond more sharply to incentives framed as losses than gains (e.g., Fryer et al., 2022; Hossain and List, 2012; Imas et al., 2016), while others fail to observe differences between the gain and loss framings (e.g., de Quidt et al., 2017; Ferraro and Tracy, 2022; Pierce et al., 2020; Camerer et al., 2016; Gneezy et al., 2017; Heffetz, 2018). In their meta-analysis of experimental studies on framing, Ferraro and Tracy (2022) calculate the average effect size of loss framing to be 0.16 standard deviations (SD) of the performance measure, the effect size being 0.33 SD for lab experiments and only 0.02 SD for field experiments, suggesting that framing effects on labor supply are highly heterogeneous and context specific.

Given the existing evidence, it is fair to say that framing effects on labor supply, while theoretically plausible, are delicate and require careful measurement which may not always be feasible in the field. Lab experiments like ours are designed to capture framing effects in strictly controlled environments. In our experiment, we intend to create framing by varying both the presence and the level of the reference piece rate (RR) across seven experimental conditions into which 249 subjects were randomly allocated. In our specific implementation, RR is the piece rate paid to the subjects at the initial, warmup stage of the experiment lasting about two minutes, after which they would work under different piece rates. The experimental conditions in which the initial RR is lower/higher than the later offered piece rates would put the subjects in a gain/loss domain in terms of the implied income (=piece rate times labor supply), whereas the conditions with no RR would serve as domain-neutral controls. This design enables the measurement of framing effects on labor supply by comparing labor supply reactions to a given change in the piece rate across the conditions in which this change would constitute a gain, loss, or be domain-neutral.

Observing independent labor supply decisions by the same individual at different piece rates is crucial for capturing individual labor supply reactions to piece rate changes, with or without framing. Outside the lab, employers can pay only one piece rate per work session. In our lab experiment, we overcome this limitation by eliciting labor supply decisions at different piece rates by the same subject and at the same time using the strategy method (Selten, 1967), which involves collecting binding labor supply decisions at each possible piece rate and having the subjects supply the amount of labor corresponding to a randomly chosen piece rate. The within-subject variation in labor supply in response to a given piece rate change in different conditions enables us to cleanly test theoretical predictions from a labor supply model adopted from Farber (2015) with minimal structural and data requirements.

The experimental studies closest to ours are Andersen et al. (2014) and Abeler et al. (2011). Both attempt to induce framing by manipulating subjects' position with respect to the income target, but produce conflicting results. In the first of Andersen et al. (2014) two field experiments with market vendors in India, they increased the treated vendors' hourly earnings in an expected way, by offering them additional income for helping with market surveys in over the next several days. In the second field experiment, the authors increased the treated vendors' income unexpectedly, by giving a surprise and one-off overpayment for the good they sold (Betel nuts) early in the trading day. Contrary to the income targeting argument, the unexpected additional income in the second experiment did not reduce labor supply, whereas the expectedly higher income in the first experiment increased labor supply.

In Abeler et al. (2011) lab experiment, subjects worked on a real-effort task with stochastic earnings. In one experimental condition, subjects earned either a fixed pay of 7 euros or a piece rate of 20 cents, each of those two earnings schemes being equiprobable and determined after the subjects finished the task. In the other condition, subjects earned either a fixed pay of 3 euros or a piece rate of 20 cents, again, equiprobably and determined after the work was done. The variation in the fixed pay across the conditions affected the subjects' target income while leaving the piece rate (20 cents) unchanged. Consistent with the income targeting argument, labor supply was significantly higher in the condition with the higher fixed pay, as subjects chose to work more at the same piece rate to stay close to the higher income target.

Our experimental design is alternative and complementary to that used in the above studies, and varies from them mainly in that, in our experiment, labor supply choices for the same task are made in loss as well as gain domains with respect to the implied income under RR. This allows testing more theoretical predictions, some of which are domain-specific, within the same experiment.

To preview our results, we find that the monotonicity in labor supply responses to piece rate changes, expected under the neoclassical labor supply model and observed in our control conditions, is disrupted by RR induced through treatment-specific piece rates paid at the warmup stage. Domain-specific labor supply responses to a given piece rate change are consistent with framing effects predicted by prospect theory: people work more for a higher piece rate when it is above RR, but they also work more for a lower piece rate when it is below RR. However, we also find that framing effects are sensitive to context: when piece rates below and above RR are present in the same treatment condition, labor supply response to piece rate changes is again monotonic. The fragility of framing effects is known: they are sensitive to small changes in the setting even within the same study context (e.g., Hossain and List, 2012), and the empirical evidence for them is patchy (Ferraro and Tracy, 2022).

To our knowledge, our study is the first to experimentally manipulate both the wage and RR in a piece rate setting (Chang and Gross, 2014, study a similar setting, but in the field). We believe this is a valuable contribution to the existing research on framing effects on labor supply. Piece rates are common in a variety of occupations, and their prevalence is likely to increase as the gig economy continues to grow. Our findings, obtained in this novel setting and in a carefully controlled experiment, strengthen the case for reference dependence in labor supply: the effect of a wage (piece rate) change depends on the reference point; therefore, an unexpected wage (piece rate) decrease can boost, rather than decrease, labor supply.

2. Theory and testable predictions

Assumptions. Our model is a simplified adaptation of Farber's (2015), in which we reinterpret the hourly wage as piece rate and hours worked as effort. As in his model, and following Kőszegi and Rabin's (2006) formulation of expectation-based reference points, we assume that the target income is based on the expected wage, operationalized in our study with RR which we manipulate experimentally. Hence our key assumption: i) the worker's income target is the level s/he would optimally choose to earn under a given RR and other relevant considerations. It is worth clarifying that assumption i) does not imply that RR is the only determinant of the income target or that the income target will necessarily change with RR (in which case our experiment varying RR would fail to produce an effect on labor supply).

The other assumptions are: ii) a quadratic cost of effort function; iii) a cap on labor supply; and iv) loss-averse preferences. Assumption ii) greatly simplifies the maths without losing essential detail. Assumption iii) adds realism by allowing for labor supply reactions to RR or actual piece rate changes to be mute (our student participants are likely time constrained, owing to class schedule, and changes in RR may fail to affect the income target). Assumption iv) reflects abundant empirical evidence on loss-averse preferences (e.g. Camerer, 2000; Wakker, 2010; Xie et al., 2018; see Brown et al., 2023 for the latest meta-analysis).

Given the above assumptions, we model the worker's labor supply h as the solution to his/her utility maximization problem:

$$\max_h U = \begin{cases} w\cdot h - T - \dfrac{\theta\cdot h^2}{2}, & \text{subject to } 0 \le h \le H \text{ and } w\cdot h \ge T \text{ (income at or above target)} \\ \lambda\cdot(w\cdot h - T) - \dfrac{\theta\cdot h^2}{2}, & \text{subject to } 0 \le h \le H \text{ and } w\cdot h < T \text{ (income below target)} \end{cases}$$

where w is the actual piece rate, $\theta > 0$ is the cost of effort parameter (assumption ii), H is the cap on labor supply (assumption iii), and $\lambda \ge 1$ is the loss-aversion parameter representing the extra disutility of income being in the loss domain relative to the target T (assumption iv). As per assumption i), the income target T is the gross earnings from the optimal labor supply that maximizes the net income $w_0\cdot h - \frac{\theta\cdot h^2}{2}$ under RR w_0 and the time constraint $0 \le h \le H$, which gives $h(w_0,\ H) = \min\left(\frac{w_0}{\theta}, H\right)$ and $T = \min\left(\frac{w_0^2}{\theta}, w_0\cdot H\right)$.[3]

Labor supply with and without RR. Separating the worker's utility maximization problem into two constrained problems, one for $0 \le h \le H,\ w\cdot h \ge T = \left(\frac{w_0^2}{\theta}, w_0\cdot H\right)$ and the other for $0 \le h \le H,\ w\cdot h \le T = \left(\frac{w_0^2}{\theta}, w_0\cdot H\right)$, and solving both simultaneously, gives the utility-maximizing labor supply as a function of the going piece rate w, the reference piece rate w_0, the loss aversion and cost of effort parameters, λ and θ, and the cap on labor supply, H:

$$h(w,\ w_0, \lambda,\ \theta,\ H) = \begin{cases} \min\left(\dfrac{\lambda\cdot w}{\theta}, H\right), & w \le \dfrac{w_0}{\sqrt{\lambda}} \\ \min\left(\dfrac{T}{w}, H\right), & \dfrac{w_0}{\sqrt{\lambda}} \le w \le w_0 \\ \min\left(\dfrac{w}{\theta},\ H\right) & w \ge w_0 \end{cases} \tag{1}$$

Labor supply in (1) is generally non-monotonic in piece rate because of the presence of RR ($w_0 > 0$) and loss aversion ($\lambda > 1$). If there is no RR or the worker is loss-neutral, labor supply is monotonically non-decreasing in piece rate:

$$h(w,\ \theta,\ H) = \min\left(\frac{w}{\theta},\ H\right). \tag{2}$$

To help compare labor supply with and without RR and generate testable predictions, Fig. 1 plots labor supply Eq. (1) for a loss-averse worker with RR $w_0 > 0$ in three cases that differ in the extent to which the cap restricts labor supply reactions to piece rate changes: $H \ge \sqrt{\lambda}\ \frac{w_0}{\theta}$ (case 1), $\frac{w_0}{\theta} < H < \sqrt{\lambda}\ \frac{w_0}{\theta}$ (case 2), and $0 < H \le \frac{w_0}{\theta}$ (case 3). Let us ignore case 3, in which labor supply at w_0 is already capped, and focus on cases 1 and 2 in which the cap is high enough to allow labor supply to react to piece rate changes.

In the gain domain, with $w \ge w_0$, labor supply with RR (Eq. (1)) is the same as without (Eq. (2)), monotonically increasing with piece rate in proportion $\frac{1}{\theta}$ (segment C on Fig. 1) before hitting the cap (segment D). Contrastingly, in the loss domain, when $w \le w_0$, labor supply with RR is non-monotonic and exceeds that without RR for a given piece rate. For a range of piece rate values $\sqrt{\frac{T\cdot\theta}{\lambda}} = \frac{w_0}{\sqrt{\lambda}} \le w \le w_0 = \sqrt{T\cdot\theta}$, labor supply $h(\cdot)$ is either capped at H (segment D, case 2), or, if the cap is not reached, $h(\cdot) = \frac{w_0^2}{\theta\cdot w} = \frac{T}{w}$, increasing just enough to maintain the target income level $T = \frac{w_0^2}{\theta}$ as the piece rate w falls further below RR w_0 (segment B). Working more at falling piece rates, or "income targeting", reflects a higher utility of reaching the income target than the effort costs of doing so for the piece rates not too far below w_0. As a result, more labor is supplied with than without RR for this range of piece rates.

Deeper in the loss domain, for relatively low piece rates $w \le \frac{w_0}{\sqrt{\lambda}} = \sqrt{\frac{T\cdot\theta}{\lambda}}$ (segment A), the cost of effort to maintain the income target becomes too high, and so less labor is supplied at lower piece rate, again. This does not mean that labor supply with RR converges to that without: in fact, labor supply with RR grows with piece more strongly than without, in proportion $\frac{\lambda}{\theta} > \frac{1}{\theta}$, because, with RR, there is more incentive to work at a given, low, piece rate – to reduce the disutility from falling below the target income level. Hence, more labor is supplied with than without RR.

Testable predictions. Provided that not everyone's labor supply at RR is capped (as in case 3), our model generates the following experimentally testable predictions.

Prediction 1. In the absence of RR, average labor supply increases with the piece rate level.

Prediction 2. In the presence of RR, average labor supply increases with piece rate in the gain domain ($w \ge w_0$) in the same proportion as

[3] Our specification of the utility-maximizing labor supply choice problem with loss aversion has been used before (e,g., Abeler et al., 2011; Andersen et al., 2014). Notice that, when either $\lambda = 1$ (no loss aversion) or $w_0 = 0$ (no RR), our specification reduces to the standard, target-free neoclassical labor supply choice problem.

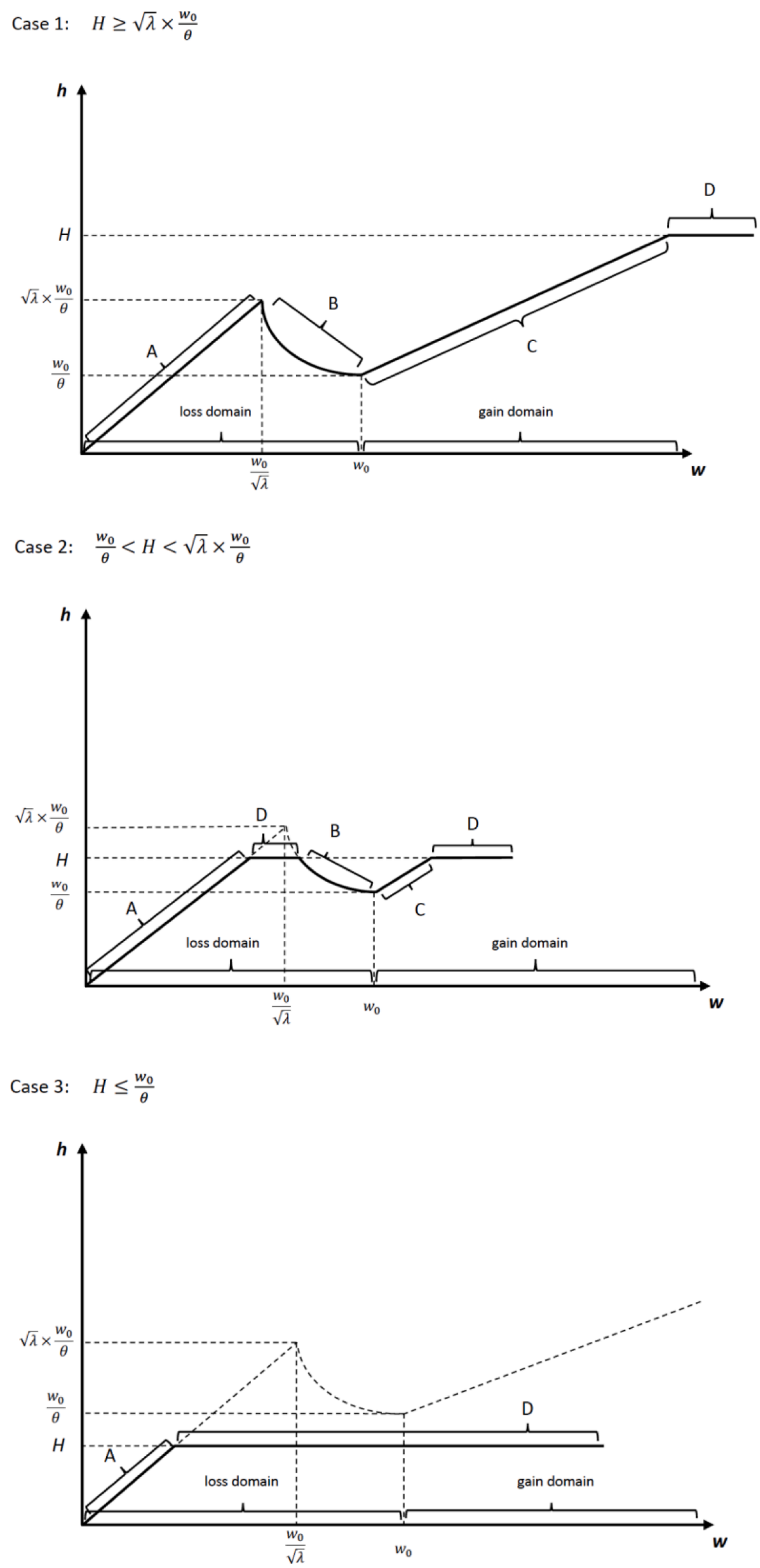

Fig. 1. Labor supply (h) as a function of piece rate (w), RR (w_0), and capped labor supply (H).

when there is no RR (Prediction 1).

Prediction 3. In the presence of RR, there will be workers whose labor supply will strictly increase as piece rate falls in the loss domain ($w \leq w_0$) so as to enable complete income targeting. (These are the workers whose degree of loss aversion, λ, satisfies $\frac{w_0}{\sqrt{\lambda}} \leq w \leq w_0$). For the rest of the workers in the loss domain, for whom $w < \frac{w_0}{\sqrt{\lambda}}$, the quantity of labor supplied may increase, decrease or stay the same as at the reference piece rate w_0, depending on the interplay between the disutility of effort and the degree of loss aversion.

Prediction 4. Given the piece rate level, labor supply is weakly higher

Table 1
Experimental conditions summary.

	Condition	Label	Piece rate in the warmup stage	Piece rates at the main stage (equiprobable, shaded areas)		
				Low	Medium	High
Control 1	No RR	X-LM	None			n/a
Trmt 1–1	Unidirectional Up	L-LM	Low			n/a
Trmt 1–2	Unidirectional Down	M-LM	Medium			n/a
Control 2	No RR	X-MH	None	n/a		
Trmt 2–1	Unidirectional Up	M-MH	Medium	n/a		
Trmt 2–2	Unidirectional Down	H-MH	High	n/a		
	Bidirectional	M-LMH	Medium			

Notes: The first letter in the condition label denotes the piece rate level offered during the warmup stage (L, M or H, or X if no piece rate was offered), and the following letters denote the combinations of possible, and equally likely, piece rates at the main stage. Conditions with a piece rate at the warmup stage are treatments, and those without are controls.

with than without RR in the loss domain.[4]

3. The experiment

To test the predictions of our model, we experimentally vary the initially paid piece rate for the real-effort task, which we intend to serve as the RR relevant for labor supply choices under future piece rates, and compare labor supply levels (=the number of tasks committed by the participants) at a given piece rate and varying RR levels.

Experimental task. We implemented the *arithmetic task,* that is, adding sets of five two-digit numbers (e.g., Niederle and Vesterlund, 2007). We chose this task because its output is mostly effort-based and little learning can take place in a short period of time (e.g., Else-Quest et al., 2010). Moreover, it is unlikely that most people derive utility from the task itself, so intrinsic motivation can be ruled out. The numbers for the arithmetic task were randomly generated on a computer, but the experiment itself was conducted using paper and pencil, and without calculators.

Piece rates. There were three piece-rate levels used in the experiment: \$0.50, \$1.00 and \$2.00 per correctly solved task (the sum of five two-digit numbers). In what follows, we refer to these as "low" (L), "medium" (M) and "high" (H) piece rates. The experiment participants were not given such labels, but were merely informed about the specific dollar amounts. We chose the piece rates to be rather generous given the time it takes to do the adding task and relative to what could be earned from part-time student jobs, to help ensure that participants in the control condition would be in the gain domain relative to the expectations-based target income they might have.

Treatment and control conditions. There were seven between-person experimental conditions that varied in the piece rate schedule offered to the participants. Each condition involved two stages: the two-minute warmup stage and the main stage, which, importantly, did not have a set time limit. The warmup stage served to familiarize the participants with the task and to help them gauge its difficulty, as well as for us to induce RR by offering piece rates at this stage in the treatment conditions. It was kept short in order to avoid tiring the participants and to minimize the potential income effect on labor supply at the main stage.

The treatment and control conditions differed by the piece rate offered during the warmup stage, and by the schedule of piece rates applicable at the main stage. The treatment conditions had a piece rate offered at the warmup stage, and the control conditions had none. Table 1 summarizes the conditions labelled according to their piece rate schedules. The first letter in the label denotes the piece rate offered during the warmup stage (L, M, H, or X if no piece rate was offered, as in the control condition), and the following letters denote the schedule of possible, and equally likely, piece rates at the main stage. For example, condition X-LM is the first control condition: it pays a flat fee of \$3 at the warmup stage, regardless of output,[5] and has the piece rate schedule L, M at the main stage. Corresponding to control condition X-LM are two treatment conditions, l-LM and M-LM, which have the same piece rate schedule at the main stage as X-LM but differ in the size of the initial piece rate: L or M. Similarly, condition X-MH is the second control condition, to which treatment conditions M-MH and H-MH correspond.[6]

Measuring labor supply. Our measure of labor supply is the number of tasks participants stated they would solve at each equiprobable piece rate at the main stage. To elicit participants' labor supply, we adopted the strategy method (Selten, 1967). The strategy method requires participants to make contingent decisions for each possible information set (Brandts and Charness, 2011). It contrasts with the direct-response method, in which subjects make a decision after receiving the pertinent information. In a comprehensive survey paper, Brandts and Charness (2011) review experimental studies that use both methods and find no treatment differences in the majority of cases. In particular, any treatment effects observed using the strategy method are also observed using the direct-response method. The advantage of the strategy method is that it provides a reasonable within-person comparison of labor supply decisions under different possible piece rate levels.

Specifically, at the beginning of the main stage, the participants were informed that there were different piece rate levels, each equally likely to be applied to calculate their earnings. They were asked to think carefully and write down the number of tasks they would commit themselves to working on under each of the equally likely piece rate levels. This enabled us to elicit labor supply decisions from each participant in an incentive compatible manner for two or three piece rates while requiring them, in the end, to work at only one.

After the participants provided their labor supply choices at each possible piece rate, the applicable piece rate level was determined for them by a random draw. To receive their payment, the participants had to correctly solve the number of tasks they had committed to under the applicable piece rate level regardless of how long it took. This is meant to motivate participants to indicate truthfully how much work they would do at each piece rate. Out of the total number of 249 participants, only two dropped out before solving the number of tasks they had specified for the randomly determined piece rate. Those participants were not paid or included in the analysis.

[4] This is similar to Abeler et al. (2011), who predict higher effort under higher target income and find supporting evidence.

[5] The flat fee amount, \$3.00, was chosen based on the average warmup stage earnings made by subjects in the treatment conditions in which a piece rate was offered at the warmup stage. The administration of this flat fee was intended to balance the wealth effects that may have been caused by the warmup stage earnings in the treatment conditions. Of course, it is possible that participants might react to the flat fee payment by calculating the implicit piece rate and focusing on that as a reference point. In that case, we would have seen no difference between the conditions with and without a piece rate at the warmup stage. In contrast, our results demonstrate a significant difference generally supportive of our theoretical predictions.

[6] In addition to the "unidirectional" conditions above, as a robustness check, we ran a bidirectional condition M-LMH in which we offered a medium reference piece rate at the warmup stage and three equiprobable piece rates at the main stage. The purpose of this additional condition was to examine whether participants would react differently when presented simultaneously with the possibility of gains and the possibility of losses rather than when shown only one or the other.

Participants and procedures. A total of 247 undergraduate students (average age 20.7 years, standard deviation 2.4 years, 57 % female) from a large Canadian university participated in the experiment and were randomly assigned to the seven conditions shown in Table 1. The conditions were randomized over morning and afternoon time slots and over days of the week.

We booked separate lab spaces for this experiment. Specifically, three small-sized breakout meeting rooms were booked as the "decision" rooms and a standard classroom down the hall was used as the "workstation" room. There was an experimenter in each of the “decision” rooms and the “workstation” room. The three “decision” rooms were used simultaneously during the experimental sessions and the seven experimental conditions were randomized to be administered across these “decision” rooms. In order to avoid peer effects and information spillovers, participants were scheduled to arrive at a given “decision” room one at a time and at least 20 min apart from any other participant assigned to that room. Thus, there was never more than one participant in each "decision" room at a time. During that 20-minute period, the participant would sign the consent form, read the instructions, complete the two-minute warm-up round (the performance in the two-minute warm-up round was graded by the experimenter in the room when the participant was reading the instructions for the experimental round), read the instructions for the experimental round and decide on their labor supply for each piece rate in the applicable piece rate schedule.[7] Then, the experimenter in the “decision” room gave the participant a die to throw to determine which piece rate would be implemented. The die-throw outcome was then recorded on the decision form by the experimenter who circled the piece rate level that was to be implemented. At that point, the participant would be escorted to the "workstation" room, where they fulfilled their work commitments. In the “workstation” room, participants could see others entering and leaving but were not allowed to communicate with them, and no interaction between participants was detected. The “workstation” room was a classroom that could accommodate to up 95 students with stationery seating. Thus, it was always sparsely occupied during the experiment. When participants were escorted to the “workstation” room, they were seated far apart from other participants who were working in the room already. Another experimenter, who was the grader and was always stationed in the workstation room, would grade the work. The grader was seated at the back of the classroom while the experimenter was at the podium/front of the classroom. If there were mistakes in the submitted work, they would need to be corrected by the participant. The interaction between the grader and a given participant took place at the seat of the participant with no verbal communication as the instructions and the grading result were presented to the participant in writing. The participant then completed a post-experiment survey that asked for their age, gender and academic major at the university, privately received their payment, and left the premises. Each participant took part in the experiment only once. In this set-up, participants had complete control over how much labor to supply at each possible piece rate, and the alternative of leisure, operationalized by permitting the participant to leave the experiment as soon as their work was done.

4. Results

Our key results, described below, are easily gleaned from simple descriptive statistics.[8] We then go beyond descriptive statistics and present structural parameter estimates of the theoretical labor supply Eq. (1).

Descriptive statistics. Fig. 2 illustrates labor supply dynamics at different piece rate levels, and Table 2 presents the corresponding averages and standard deviations. While average labor supply at the warmup stage varies little by treatment, its variation at the main stage is stronger and is mostly consistent with our model predictions.[9]

Result 1. On average, in the absence of a reference piece rate at the warmup stage, people work more for a higher piece rate than for a lower one.

In control condition X-LM, labor supply was 3.66 higher for the medium piece rate than for the low piece rate ($t(28) = 3.31, p < 0.01$). In the second control condition, X-MH, the increase in labor supply in response to higher piece rate is less sharp, 1.87 ($t(30) = 1.91, p = 0.066$), but is directionally the same as in X-LM and is not significantly different in magnitude (3.66 vs. 1.87, $t(59)$=1.21, $p = 0.231$). Higher labor supply at higher piece rates in both conditions supports Prediction 1 and implies that in the absence of RR the substitution effect in labor supply dominates the income effect – at least in the short run and within our experimental setting.

Result 2. On average, people work more for a higher piece rate in the gain domain (i.e., when the piece rate is above RR).

In the treatment conditions l-LM and M-MH, the piece rates at the warmup stage were low and medium, respectively. Then, at the main stage, participants were asked how much they would work for a higher piece rate as well as for the previously experienced RR. Specifically, the two possible piece rates were low and medium in condition l-LM and medium and high in condition M-MH. The within-person differences in labor supply at the higher and lower piece rates at the main stage are 3.18 and 5.92 in treatments l-LM and M-MH, respectively, both significantly different from zero ($t(38) = 3.66, p < 0.01$, and $t(37) = 5.69, p < 0.01$ respectively). The willingness to work more when piece rates increase from the previously experienced RR supports Prediction 2.

Prediction 2 also states that, in the gain domain, labor supply increases with piece rate in the same proportion as without RR. While labor supply reactions to piece rate changes in conditions X-LM and l-LM are comparable, conditions X-MH and M-MH produce significantly different results: 1.87 vs. 5.92 ($t(68) = 4.39, p < 0.01$). Still, labor supply positively reacts to piece rate increases in both cases.

Result 3. On average, people work more for a lower piece rate in the loss domain (i.e., when the piece rate is below RR).

In the treatment conditions M-LM and H-MH, the piece rates at the warmup stage were medium and high, respectively. At the main stage, participants were asked how much they would work for a lower piece rate as well as for the previously experienced piece rate meant to serve as RR. Specifically, the two possible piece rates were low and medium in condition M-LM and medium and high in condition H-MH. The within-person differences in labor supply at the higher and lower piece rates are -5.00 and -4.36 in condition M-LM and H-MH, respectively, both statistically significant ($t(40) = 4.22$, $p < 0.01$, and $t(38) = 3.91$, $p < 0.01$ respectively). The willingness to work significantly more when piece rates decrease from the previously experienced RR supports Prediction 3.

The interpretation of Results 2 and 3 based on our model is that there is a kink in the labor supply curve at the previously experienced piece rate that serves as RR. Piece rates falling below RR trigger income targeting, as people try to mitigate the disutility from earning less than the

[7] All experimental materials, i.e., consent form, task instructions, warm-up round workbook, instructions for the experimental round, decision form, experimental round workbook, and post-experiment questionnaire, were presented to each participant one at a time.

[8] Participants' age, gender, and major of study are balanced across the conditions and do not have significant impact on participants' labor supply response to piece rate changes.

[9] Our results are not driven by outliers. We have checked this by rerunning the analysis reported in Table 2 on the sample without the top and bottom 5% of labor supply choices. The results are very similar to those obtained on the whole dataset.

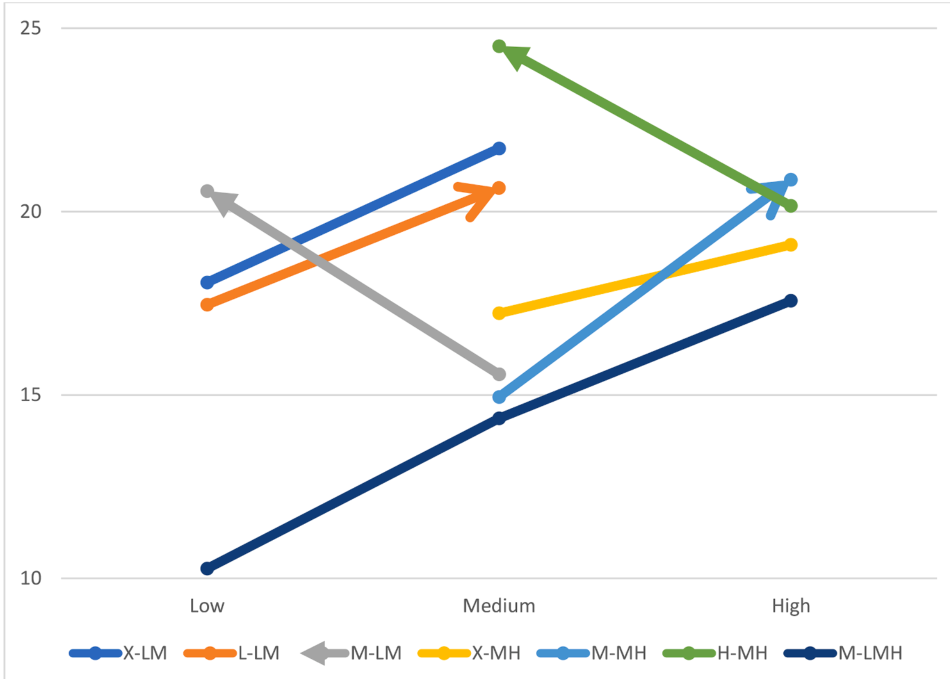

Fig. 2. Mean labor supply by experimental condition and piece rate level.

Table 2
Mean labor supply by condition and piece rate level.

	Condition	Warmup	Main stage, by piece rate level			Changes in labor supply with piece rate	
		Stage	Low	Medium	High	L to M	M to H
Control 1	X-LM	2.552	18.069	21.724		3.655**	
	$n = 29$	(0.270)	(2.153)	(2.207)		(1.106)	
Trmt 1–1	L-LM	2.744	17.462	20.641		3.179**	
	$n = 39$	(0.274)	(1.794)	(1.865)		(0.868)	
Trmt 2–1	M-LM	2.439	20.561	15.561		−5.000**	
	$n = 41$	(0.204)	(2.960)	(2.126)		(1.184)	
Control 2	X-MH	2.839		17.226	19.097		1.871
	$n = 31$	(0.275)		(2.395)	(2.319)		(0.980)
Trmt 1–2	M-MH	3.026		14.947	20.868		5.921**
	$n = 38$	(0.198)		(1.773)	(2.224)		(1.041)
Trmt 2–2	H-MH	3.077		24.513	20.154		−4.359**
	$n = 39$	(0.253)		(1.893)	(1.637)		(1.115)
Bidirectional	M-LMH	2.533	10.267	14.367	17.567	4.100**	3.200**
	$n = 30$	(0.218)	(1.493)	(1.808)	(2.022)	(0.782)	(0.839)
Average across conditions		2.753	16.950	18.470	19.551	1.065	1.514
p-value test of equality across conditions		0.359	0.018	0.001	0.704	0.000	0.000

Notes: The first letter in the condition label denotes the piece rate level offered during the warmup stage (L, M or H, or X if no piece rate was offered), and the following letters denote the combinations of possible, and equally likely, piece rates at the main stage. Conditions with a piece rate at the warmup stage are treatments, and those without are controls. Standard deviations are in parentheses. (**) [*] indicates changes are significant at the (1 %) [5 %] level in a two-sided *t*-test with regression standard errors clustered by participant ID.

target income by working more. This behavior stems from loss aversion: the utility from lower financial losses relative to the reference point is traded for disutility from extra work effort. On the other hand, in the gain domain, when piece rates rise above RR, loss aversion no longer affects effort and substitution effects dominate income effects, just like in the neoclassical labor supply model (Eq. (2)).

Result 4. On average, at a given piece rate, people work more in the loss domain than in the gain domain.

We now turn from within- to between-participant, cross-condition comparisons. Starting with labor supply at the low piece rate, we find that, in condition M-LM, in which earnings at the low piece rate would be in the loss domain, average labor supply (20.56) is significantly higher than that in the loss-free conditions X-LM and l-LM featuring the same menu of piece rates (average 17.72, ($t(108)$= 3.10, $p < 0.01$). A similar exercise for labor supply under the medium piece rate reveals that the average labor supply in condition H-MH when the income under the medium piece rate is in the loss domain (24.513), is higher than the average labor supply in the loss-free but otherwise identical conditions X-MH and M-MH (average 15.97), also a statistically significant difference ($t(110)$= 3.34, $p < 0.01$). Labor supply at a given piece rate being higher when the income it brings is in the loss domain supports Prediction 4.

Result 5. Income targeting is present but incomplete.

Although there is income targeting in the loss domain (people working more for a lower piece rate to bring income closer to the target, Result 3), income nonetheless drops as the piece rate falls. Table 3

Table 3
Average earnings at the main stage in the unidirectional-down conditions.

Condition	Low piece rate ($0.50)	Medium piece rate ($1.00)	High piece rate ($2.00)
M-LM, $n =$ 41	$10.28	$15.56	
H-MH, $n =$ 39		$24.51	$40.30

presents the average earnings that would have occurred at each piece rate at the main stage based on the labor supply commitments in the two treatment conditions, M-LM and H-MH, in which income targeting occurs. In these conditions, the dominance of loss-aversion over substitution effects in the loss domain causes an increase in labor supply that partially offsets the drop in income but does not eliminate it. The difference in income earned given labor supply commitments under the two potential piece rates is still both economically and statistically significant, indicating that income targeting is not complete in either condition ($t(40) = 4.32$, $p < 0.01$ and $t(38) = 7.83$, $p < 0.01$ respectively).

There are two possible reasons for incomplete income targeting: i) the effort required to meet the income target given the piece rate is too costly, and/or ii) the labor supply is capped. Both appear to be empirically relevant. Starting with the cap, Table 4 reports categorized labor supply responses to piece rate changes by condition. In all conditions, a large share of participants (between 26.7 % and 64.5 %) offered the same labor supply at all piece rate levels, which is consistent with their labor supply having been capped. This is true whether or not there was an initial piece rate at the warm-up stage. In the bidirectional condition M-LMH, 26.7 % appear to be completely capped with labor supply unchanged at all three piece-rate levels, while an additional 26.7 % are constrained by the cap only at the medium and high piece rates. In no case was labor supply unchanged at the low versus the medium, but increased at the high piece rate, which would have been inconsistent with the cap on labor supply. Thus, a detailed analysis of individual responses suggests the presence of a cap on labor supply.

Table 4
Individual labor supply responses to piece rate changes, categorized.

	Condition	The same for all piece rate levels	More under higher piece rate	More under lower piece rate	Complete Income Targeting	Partial Income Targeting
Control 1	X-LM	17	10	2		
	$n =$ 29	(58.6 %)	(34.5 %)	(6.9 %)		
Trmt 1–1	L-LM	18	19	2		
	$n =$ 39	(46.2 %)	(48.7 %)	(5.1 %)		
Trmt 2–1	M-LM	17	2	9 + 13	9	13
	$n =$ 41	(41.5 %)	(4.9 %)	(53.7 %)	(22.0 %)	(31.7 %)
Control 2	X-MH	20	8	3		
	$n =$ 31	(64.5 %)	(25.8 %)	(9.7 %)		
Trmt 1–2	M-MH	11	26	1		
	$n =$ 38	(28.9 %)	(68.4 %)	(2.6 %)		
Trmt 2–2	H-MH	13	4	6 + 16	6	16
	$n =$ 39	(33.3 %)	(10.3 %)	(56.4 %)	(15.4 %)	(41.0 %)
Bidirectional	M-LMH	8	(8 *+14**)	0		
	$n =$ 30	(26.7 %)	(73.3 %)	(0 %)		

Notes: The counts in the table show the number of participants whose labor response belongs to a given category, and the percentages in parentheses represent their share in the corresponding condition. For instance, 17 people chose to supply the same amount of labor under the low and medium piece rates in condition X-LM, which is 58.6 % of the total of 29 participants in that treatment.

* Unchanged between high and medium piece rate, but less under low piece rate,.

** More under high than under medium piece rate, and more under medium piece rate than under low piece rate.

Turning to the remaining participants, in the control conditions without RR, the overwhelming majority offered more under the higher piece rate. The same behavior prevails in the conditions with RR, in the gain domain. In sharp contrast, and consistent with Prediction 3, in the loss domain, the majority of the participants, though not all, worked more under the lower piece rate. Focussing on the unidirectional treatments with RR (we deal with the bi-directional M-LMH next), of the total of 44 participants who chose to work more under the lower piece rate in the loss domain, 15 increased their labor supply by just enough to keep their income constant (thus achieving complete income targeting, depicted in segment B on Fig. 1), while 29 did not increase their labor supply enough to keep their income constant (incomplete income targeting). The latter behavior is consistent with being on the upward sloping segment A of the theoretical labor supply line (Fig. 1) to the left of the downward sloping, income-targeting, segment B or on the capped labor supply segment D, but above the labor supply at the reference piece rate in both cases. The remaining 6 participants who offered more under the higher piece rate in those conditions acted consistently with being on the leftmost segment A of the supply line below the labor supply level at the reference piece rate. Thus, costly effort and capped labor supply are both empirically relevant explanations for the incomplete income targeting we see in the data.

Result 6. On average, in the bidirectional condition M-LMH, people work more for a higher piece rate across all piece rate levels.

In condition M-LMH, participants received the medium piece rate at the warmup stage, followed by equiprobable low, medium, or high piece rates at the main stage. This condition was designed as a robustness check to examine whether the apparent kink in the labor supply schedule that emerged when comparing the unidirectional-up and -down conditions would continue to manifest itself in a bidirectional condition, which could result in either gain or loss relative to RR and the implied target income. It did not. Labor supply increased by 4.10 ($p < 0.001$) moving from the low to the medium piece rate and by a further 3.20 ($p < 0.001$) moving from the medium to the high piece rate, indicating the dominance of substitution over income effects in both gain and loss domains. While labor supply increases in response to higher piece rates are similar in magnitude to those observed in the gain domain in the other conditions, the absence of the reverse effect in the loss domain is inconsistent with the results from the other conditions, and is therefore puzzling. We discuss possible explanations in the concluding section.

Structural estimation results. While descriptive statistics directionally support most of our predictions in most of the conditions, they do not reveal the values of the structural parameters of the model that generates these predictions: the labor supply cap (H), loss aversion (λ), or the difficulty of effort (θ). As well as being interesting on its own and helpful for comparing our results with other studies, structural estimation of (1) is also instructive for understanding the identifiability issues accompanying the labor-supply effects of interest, and how these issues can be addressed in our or similarly designed experiments.

To bring Eq. (1) to an estimable form given our data, we take within-participant differences in labor supply (Δh) at the higher (w_{higher}) and lower (w_{lower}) piece rates applying in each condition. The resulting expressions, their expected signs, the requirements for them to hold, and the data sources available to estimate them are presented below.

$\Delta h = h\left(w_{higher}\right) - h(w_{lower})$

Expression	Observable condition	Data source
$\min\left(\frac{w_{higher}}{\theta}, H\right) - \min\left(\frac{\lambda \cdot w_{lower}}{\theta}, H\right) ?0$	Loss domain andincomplete income targeting	Conditions M – LM, H – MH, (M – LMH)
$= \min\left(\frac{w_{higher}}{\theta}, H\right) - \min\left(\frac{\min\left(\frac{w_{higher}^2}{\theta}, w_{higher}H\right)}{w_{lower}}, H\right) \leq 0$	Loss domain andcomplete income targeting	M – LM, H – MH, (M – LMH)
$\min\left(\frac{w_{higher}}{\theta}, H\right) - \min\left(\frac{w_{lower}}{\theta}, H\right) \geq 0$	Gain domain orno reference wage	X – LM, L – LM, X – MH, M – MH, (M – LMH)

The above expressions clarify the problem of identifiability of the effects of piece rate changes on labor supply. The directions of the predicted effects rule out a positive result in the middle expression (more labor supply at RR than at the lower one with complete income targeting) and a negative result in the bottom expression (less labor supply at higher piece rate in the gain domain). They are, however, ambiguous in the top expression because of the cusp in the labor supply schedule at $w = \frac{w_0}{\sqrt{\lambda}}$ for $\lambda > 1$. Structural estimation is a valuable complementary testing strategy. It is feasible because our experiment has generated all the data necessary to identify the model parameters.

Table 5 reports parameter estimates of Eq. (1) obtained from various specifications (technical detail in the notes to the table). Consistent with our earlier observation of many participants not changing their labor supply in response to piece rate changes (Table 4), there is a cap on labor supply averaging at about 18, which corresponds to the 56th percentile in the actual labor supply distribution. The loss aversion estimates range from 2.3 to 3, which is broadly consistent with the estimates reported in the existing literature (about 2, Brown et al., 2023), and are all significantly different from 1, implying income targeting at piece rates down to $\frac{1}{\sqrt{3}} \approx 0.58$ of RR (as long as the cap does not bind). Ignoring the cap produces lower estimates of loss aversion, as muted labor supply response is attributed to weaker loss aversion rather than to the cap. Excluding the observations from condition M-LMH produces higher estimates of loss aversion, which is expected since we did not observe income targeting in that condition – for the reasons we speculate in conclusion but cannot test. All in all, the structural estimation results confirm our theory predictions, are broadly consistent with descriptive evidence, and enrich our understanding of labor supply in the gain and loss domains by explicitly accounting for the labor supply cap.

Table 5
Structural estimates of the theoretical labor supply Eq. (1).

Parameter	(1)	(2)	(3)	(4)	(5)
Loss aversion λ	2.382	2.341	2.618	2.501	3.003
	(0.190)	(0.187)	(0.189)	(0.372)	(1.076)
Difficulty of effort θ	0.202	0.209	0.189	0.318	0.312
	(0.020)	(0.022)	(0.021)	(0.046)	(0.056)
Cap on labor supply H	ignored	ignored	ignored	17.726	17.669
				(3.379)	(4.393)
Data sample	all	all	no M-LMH	all	no M-LMH
Controls included?	no	yes	yes	yes	yes

Notes: The structural parameters are estimated with nonlinear least squares method using Stata **nl** command (script available). The regression equation underlying specifications in Columns (4) and (5) is $\Delta h = [(\frac{w_{higher}}{\theta}, H) - (\frac{w_{lower}}{\theta}, H)]\cdot(1 - I(loss)) + \left[(\frac{w_{higher}}{\theta}, H) - \left(\frac{\left(\frac{w_{higher}^2}{\theta}, w_{higher}H\right)}{w_{lower}}, H\right)\right]\cdot I(loss)\cdot I(complete) + [(\frac{\lambda \cdot w_{lower}}{\theta}, H)]\cdot I(loss)\cdot(1 - I(complete)) + controls + error$, where $I(loss) = 1$ if income is in the loss domain and 0 if income is in the gain domain and $I(complete) = 1$ if income targeting is complete and 0 otherwise. The specifications in Columns (1)-(3) ignore the cap on labor supply and are based on the simplified regression (think of the above regression with $H \to \infty$): $\Delta h = (w_{higher} - w_{lower})\frac{1}{\theta}\cdot(1 - I(loss)) + \left(\frac{w_{higher}}{\theta} - \frac{w_{higher}^2}{\theta \cdot w_{lower}}\right)\cdot I(loss)\cdot I(complete) + \left(\frac{w_{higher}}{\theta} - \frac{\lambda \cdot w_{lower}}{\theta}\right)\cdot I(loss)\cdot(1 - I(complete)) + controls + error$. Since the indicators $I(loss)$, $I(complete)$ are random variables estimated from the experimental results, the estimates are bootstrapped with 200 repetitions. Controls comprise gender, age, and study major.

5. Discussion and conclusion

How labor supply reacts to changes in the price of labor is among the fundamental questions of labor economics. As the world of work moves away from a rigid nine-to-five schedule with fixed wages to more flexible arrangements, such as gig employment where both hours and piece rates fluctuate, this question becomes ever more important. Prospect theory predicts framing effects of reference income: people would work more for a lower wage to meet the income target. However, while some research has produced findings consistent with this prediction (e.g., Camerer et al., 1997; Fehr and Goette, 2007; Abeler et al., 2011; Crawford and Meng, 2011; Andersen et al., 2014), there are also studies that fail to find supporting evidence (Farber, 2005, 2008, 2015). Framing effects of reference income on labor supply are known to be fragile and context-specific (Ferraro and Tracy, 2022). The contribution of our study to the existing literature is to carefully measure these framing effects in a lab experiment, and in the specific context in which the reference income is induced by reference piece rate (RR).

Why is this a valuable contribution? Given the empirically observed heterogeneity of framing effects, it is useful to find ways to better capture the relevant income target by linking it to the context in which it might emerge. Reference piece rate are an intuitively appealing contextual factor whose relevance is suggested not only by our adaptation of Farber's (2015) theoretical model but also by empirical evidence on labor supply effects of pay transparency (Liu-Kiel et al., 2013; Breza et al., 2018).

Let us take stock of our results to assess the evidence for framing effects via RR. We find that people work more for higher piece rates above RR (Result 2) and also more for lower piece rates below RR (Result 3), which is consistent with targeting the RR-induced income level (Result 4), albeit incomplete (Result 5), but is in stark contrast to the monotonic effect of piece rate on labor supply in RR-free conditions (Result 1). Confusingly, when piece rate can go above and below RR, labor supply is again monotonically increasing in piece rate (Result 6).

With all our results except one (Result 6) supporting our model predictions, to what extent can we argue that RR-induced framing is among the factors affecting people's labor supply at a given piece rate level? Results 1 and 2 do not contradict framing but they could have occurred even in the absence of it and Result 6 is inconsistent with framing.[10] Of the remaining three results, the strongest support for framing comes from comparing the outcomes of the pairs of unidirectional conditions with the same sets of possible piece rates but different RR levels, namely conditions l-LM vs. M-LM and M-MH vs. H-MH (Result 3). If framing through RR did not matter, we would observe the same labor supply reactions to a given piece rate change (say, from L=\$0.50 to M=\$1.00) across the conditions. Yet, people *decrease* labor supply when M changes to L if RR is at L, so there is no loss at either piece rate, but *increase* labor supply in response to the same piece rate change if RR is at M, so there is loss at L. We are not aware of a theoretical mechanism other than framing that would lend an equally direct rationalization to this result.

Of course, framing is not the only force affecting labor supply choices in the presence of reference points, as it seems to apply in some contexts but not in others. In bidirectional condition M-LMH, where piece rates may either rise or fall relative to RR, labor supply monotonically increases with piece rates (Result 6), in contrast with Result 3 obtained under the unidirectional conditions. One possible explanation for this puzzle is a preference for consistency (Cialdini et al., 1995; Guadagno and Cialdini, 2010; Falk and Zimmermann, 2013), that is, a desire to "respond to … situations in a manner consistent with prior attitudes, behaviors, and commitments, particularly when … consistency is salient to them" (Guadagno and Cialdini, 2010, p. 153). A nonmonotonic labor-supply response to piece rate changes that loss aversion alone would predict in condition M-LMH may have appeared less consistent to the participants than the monotonic responses in the other, unidirectional, treatments, causing them to align their choices towards more consistency.

Another possible explanation is provided by the salience theory of choice under risk (Bordalo et al., 2012): more alternatives in condition M-LMH than in the other conditions make the loss less likely to occur (1/3 of the time, rather than ½), and thus less salient, shifting the focus from offsetting the loss by working more to substitution between work and leisure in response to changing piece rates as observed in the treatments with no reference piece rate. Again, framing is an elusive and context-specific phenomenon; for instance, Hossain and List (2012) find framing to exist in team decisions but not individual ones.

In addition to the puzzling disappearance of framing effects in the bidirectional condition M-LMH, which we cannot explain empirically within the confines of our experimental design, our study has further limitations. First, our 247 data points in seven conditions is not a large number of observations per condition. The timing of our experiment, which was carried out in early March 2020, severely limited the amount of data we could collect as the university was shut down for all in-person activities as of March 13 2020, as an emergency response to the Covid-19 pandemic. Nonetheless, our statistical tests had sufficient power to reject the null hypothesis and establish statistical significance for all questions of interest. Second, there is an issue of external validity. This is a common concern in experimental research, and one could rightly question the applicability of our findings to the "real" working environment. However, the size and salience of the rewards we offered, the earnings the subjects made relative to a typical student's income, and the magnitude of the variation in their labor supply choices across the experimental conditions all suggest that we observed economic behaviors whose logic was (to a large extent) consistent with our theoretical priors and earlier empirical findings from diverse contexts, including the field. Third, the RR offered at the warmup stage could produce an income effect on labor supply at the main experimental stage. While this limitation is important to acknowledge, the short duration of the warmup stage makes the potential income effect unlikely to be decisive.

To conclude, a plausible interpretation of our results is that RR affects the income target that matters for labor supply decisions, at least in some settings, by causing people who are loss-averse to work more in order to mitigate the financial loss when piece rates fall below RR. However, RR-induced income targets are not omnipresent and their salience may vary depending on the earnings history and the context in which labor supply decisions are made. For instance, people earning higher piece rates in the past will not always work more at lower piece rates, since they may find themselves below the piece rate range consistent with income targeting. Further research should look more closely at the link between RR and income targets and consider a wider spectrum of behavioural aspects of labor-supply decision making, including preferences other than loss aversion.

CRediT authorship contribution statement

C. Bram Cadsby: Conceptualization, Investigation, Writing – original draft. **Fei Song:** Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Nick Zubanov:** Conceptualization, Data curation, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

Appendix. Experimental instructions

Instructions

Thank you for participating today.

All of your responses in this study will remain completely anonymous. It is important that during this experiment you do not talk or make any noise that might disrupt others around you. If you have any questions, please raise your hand and the experimenter will answer your questions individually.

During this experiment you will be asked to add up sets of five double-digit integers such as the following.

[10] Taken individually, Result 6 coming from condition M-LMH is not necessarily inconsistent with our model. Combinations of the loss-aversion and cost-of-effort parameters exist under which labor supply in the loss domain will be below that under RR, resulting in the observed labor supply monotonically increasing with piece rate across domains, as in condition M-LMH. However, the inconsistency of the results in conditions M-LMH (higher labor supply under medium than low piece rate) versus M-LM (higher labor supply under low than medium piece rate) combined with random assignment of participants into conditions and the balance of their observable characteristics across treatments make this explanation unlikely.

98	42	69	50	78

The first round is a 2-minute warm-up round for you to get familiar with the task. It will be followed by an experimental round. Both the warm-up and the experimental rounds will be used to calculate your earnings as explained below. You are not allowed to use a calculator, but may write numbers down on scratch paper provided by us. The numbers are randomly drawn and each problem is presented as above.

You will have a *Workbook* that will contain all of your work. Your task in each round is to solve such summing problems. Your earnings in this experiment will depend on your performance and/or the specific compensation method applied to the warm-up and the experimental rounds. Once we begin the experiment, you will not be able to look ahead to future pages or to go back to previous pages.

To ensure confidentiality, just write down your participant number on each page of the *Workbook*. Please do not write your name on any of these materials.

Please make sure that you completely understand the instructions for the experiment. Once again, remember not to make any noises that might disturb others around you. If you have any questions, raise your hand and we will answer your questions individually.

Warm-up Round:

Please write all of your work in this *Workbook* and turn pages only when instructed to do so.

The next round is a warm-up round and it will last for 2 min.

Please solve the problems in the order presented (i.e. You are NOT allowed to skip problems).

You will earn $1.00 for each problem you solve correctly.

Thus, your total earnings for this round will be: $1.00 × the number of problems solved correctly.

Sample main stage:

For this stage, there are 2 possible pay levels below:

Method A	**Method B**
$0.50/problem solved	$1.00/problem solved

Each of these two options has an equal chance of being chosen for your actual work output and payment. You will roll a 6-sided die. If the die-roll outcome is an odd number (i.e. 1, 3 or 5), then Method A will be chosen for both work output and payment. If the die-roll outcome is an even number (i.e. 2, 4 or 6), then Method B will be chosen for both work output and payment.

Please write down the number of problems you'd like to commit yourself to work on under each method. Please note that these are binding commitments. When one of these two methods is chosen at random, you are committed to solve the number of problems you committed yourself to solve in order to receive payment. They must be solved correctly. If there are any incorrect answers, you will be given the opportunity to redo those problems. You will then be paid according to the payment method randomly chosen by the die roll.

Now please take a minute to make your decisions on how many problems you'd like to solve under each method:

Method A	**Method B**
$0.50/problem solved	$1.00/problem solved
I will solve ______problems	**______problems**

References

Abeler, J., Falk, A., Huffman, D., Goette, L., 2011. Reference point and effort provision. American Econ. Rev.w 101, 470–492.

Andersen, S., Brandon, A., Gneezy, U., and List, J. (2014). Toward an understanding of reference-dependent labor supply: theory and evidence from a field experiment. *NBER Working Papers 20695*.

Bordalo, P., Gennaioli, N., Shleifer, A., 2012. Salience theory of choice under risk. Quart. J. Econ. 127 (3), 1243–1285.

Brandts, J., Charness, G., 2011. The strategy versus the direct-response method: a first survey of the experimental comparisons. Experim. Econ. 375–398.

Breza, E., Kaur, S., Samsadani, Y., 2018. The morale effects of pay inequality. Quart. J. Econ. 133 (2), 611–663.

Brown, A.L., Imai, T., Vieider, F., Camerer, C., 2023. Meta-analysis of empirical estimates of loss-aversion. J. Econ. Lit forthcoming.

Brownback, A., Sadoff, S., 2020. Improving college instruction through incentives. J. Political Econ. 128 (8), 2925–2972.

Camerer, C., 2000. Prospect theory in the wild: evidence from the field. In: Kahneman, D., Tversky, A. (Eds.), Choices, Values, and Frames. Cambridge University Press.

Camerer, C., Babcock, L., Loewenstein, G., Thaler, R., 1997. Labor supply of New York city cabdrivers: one day at a time. Quart. J. Econ. 112 (2), 407–441.

Camerer, C.F., 2004. Prospect theory in the wild: evidence from the field. In: Kahneman, D., Tversky, A. (Eds.), Advances in Behavioral Economics. Cambridge University Press, Cambridge, UK, pp. 148–161.

Camerer, C.F., Dreber, A., Forsell, E., Ho, T.H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., Wu, H., 2016. Evaluating replicability of laboratory experiments in economics. Science 351 (6280), 1433–1436.

Carrington, W., 1996. The Alaskan labor market during the pipeline era. J. Political Econ. 104, 186–218.

Chang, T., Gross, T., 2014. How many pears would a pear packer pack if a pear packer could pack pears at quasi-exogenously varying piece rates? J. Econ. Behav. Organiz. 99, 1–17.

Chou, Y.K., 2002. Testing alternative models of labour supply: evidence from taxi drivers in Singapore. Singapore Econ. Rev. 47 (1), 17–47.

Cialdini, R.N., Trost, M.R., Newsom, J.T., 1995. Preference for consistency: the development of a valid measure and the discovery of surprising behavioral implications. J. Pers. Soc. Psychol 69 (2), 318–328.

Crawford, V., Meng, J., 2011. New York City cabdrivers' labor supply revisited: reference-dependent preferences with rational-expectations targets for hours and income. American Econ. Rev. 101, 1912–1932.

de Quidt, J., Fallucchi, F., K.lle, F., Nosenzo, D., Quercia, S., 2017. Bonus versus penalty: how robust are the effects of contract framing? J. Econ. Sci. Assoc. 3 (2), 174–182.

DellaVigna, S., Lindner, A., Reizer, B., Schmieder, J., 2017. Reference-dependent job search: evidence from Hungary. Q. J. Econ 132, 1969–2018.

Else-Quest, N., Hyde, H., Linn, M., 2010. Cross-national patters of gender differences in mathematics: a meta-analysis. Psychol. Bull 136, 103–127.

Ericson, K.M.M., Fuster, A., 2011. Expectations as endowments: evidence on reference-dependent preferences from exchange and valuation experiments. Quart. J. Econ. 126, 1879–1907.

Falk, A., Zimmermann, F., 2013. A taste for consistency and survey response behavior. CESifo Econ. Stud 59 (1), 181–193.

Farber, H.S., 2005. Is tomorrow another day? The labor supply of New York City cabdrivers. J. Political Econ. 113 (1), 46–82.

Farber, H.S., 2008. Reference-dependent preferences and labor supply: the case of New York City taxi drivers. American Econ. Rev. 98 (3), 1069–1082.

Farber, H.S., 2015. Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. Quart. J. Econ. 130 (4), 1975–2026.

Fehr, E., Goette, L., 2007. Do workers work more if wages are high? Evidence from a randomized field experiment. Amer. Econ. Rev. 97, 298–317.

Ferraro, P.J., Tracy, J.D., 2022. A reassessment of the potential for loss-framed incentive contracts to increase productivity: a meta-analysis and a real-effort experiment. Experimen. Econ. 25, 1441–1466.

Fryer Jr., R.G., List, J., Sadoff, D., 2022. Enhancing the efficacy of teacher incentives through framing: A field experiment. Am. Econ. J.: Econ. Policy 14 (4), 269–299.

Ganegoda, D.B., Folger, R., 2015. Framing effects in justice perceptions: prospect theory and counterfactuals. Organ Behav. Hum. Decis. Process 126, 27–36.

Gneezy, U., Goette, L., Sprenger, C., Zimmermann, F., 2017. The limits of expectation-based reference dependence. J. Eur. Econ. Assoc 15, 861–876.

Guadagno, R.E., Cialdini, R.N., 2010. Preference for consistency and social influence: a review of current research findings. Soc. Influ 5 (3), 152–163.

Hardie, B.G., Johnson, E.J., Fader, P.S., 1993. Modeling loss aversion and reference dependence effects on brand choice. Market. Sci. 12, 378–394.

Heffetz, O., 2018. Are reference points merely lagged beliefs over probabilities? J. Econ. Behav. Organ. 181, 252–269.

Hossain, T., List, J., 2012. The behavioralist visits the factory: increasing productivity using simple framing manipulations. Manage Sci 58, 2151–2167.

Imas, A., Sadoff, S., Samek, A., 2016. Do people anticipate loss aversion? Manag. Sci. 63 (5), 1271–1284.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–291.

Kőszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Quart. J. Econ. 121, 1133–1165.

Lazear, E., 2000. Performance pay and productivity. Amer. Econ. Rev. 90 (5), 1346–1361.

Liu-Kiel, H., Cadsby, C.B., Schenk-Mathes, H., Song, F., Yang, X, 2013. A cross-cultural real-effort experiment on piece rate-inequality information and performance. BE J. Econ. Analy. Policy 13 (2), 1095–1120.

Milkman, K.L., Mazza, M.C., Shu, L., Tsay, C.J., Bazerman, M.H., 2012. Policy bundling to overcome loss aversion: a method for improving legislative outcomes. Organ Behav. Hum. Decis. Process 117, 158–167.

Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? Quart. J. Econ. 122, 1067–1101.

Oettinger, G., 1999. An empirical analysis of the daily labor supply of stadium vendors. J. Polit. Econ. 107, 360–392.

Pierce, L., Rees-Jones, A., Blank, C. (2020). The negative consequences of loss-framed performance incentives. National Bureau of Economic Research Working Paper Series https://www.nber.org/papers/w26619.Pdf.

Polman, E., 2012. Self–other decision making and loss aversion. Organ Behav. Hum. Decis. Process 119, 141–150.

Selten, R., 1967. Die strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopol experiments. In: Sauermann, H. (Ed.), Beiträge Zur Experimentellen Wirtschaftsforschung. J.C.B. Mohr (Siebeck), Tübingen, pp. 136–168.

Shearer, B., 2004. Piece rates, fixed wages and incentives: evidence from a field experiment. Rev. Econ. Stud. 71 (2), 513–534.

Stafford, T.M., 2015. What do fishermen tell us that taxi drivers do not? An empirical investigation of labor supply. J. Labor Econ 33 (3), 683–710.

Stafford, T.M. (2018). Do workers work more when earnings are high?. *IZA World of Labor Working Paper.*

Wakker, P., 2010. Prospect theory: For risk and Ambiguity. Cambridge University Press.

Xie, Y., Hwang, S., Pantelous, A.A., 2018. Loss aversion around the world: empirical evidence from pension funds. J. Bank. Finance 88, 52–62.