THE LIMITS OF EXPECTATIONS-BASED REFERENCE DEPENDENCE

Uri Gneezy
University of California San Diego

Lorenz Goette
University of Bonn

Charles Sprenger
University of California San Diego

Florian Zimmermann
University of Zurich

Abstract
Theories of expectations-based reference-dependent preferences have provided a critical modeling innovation, incorporating a structured theory of the formation of reference points. An important prediction of these models is a monotone response in behavior to changes in expectations. To test such models we conduct a real-effort experiment manipulating expectations and examining consequences on effort provision. In contrast to the theory, we document substantial nonmonotonicities in the effort response to changing expectations. Our results provide some evidence on the limitations of expectations-based reference dependence. (JEL: D81, D84, D12, D03)

1. Introduction

Reference-dependent preferences are at the heart of many behavioral theories. Pioneered by Kahneman and Tversky (1979), reference-dependent preferences have found their way into many economic applications. A key model feature—loss aversion around a reference point—can rationalize behaviors that are at odds with the implications of standard expected utility maximization. Applications range from explaining why investors tend to hang on too long to losing stocks (Shefrin and Statman 1985; Odean 1998), or to houses that they would sell at a loss (Genesove and Mayer 2001), the endowment effect (Kahneman, Knetsch, and Thaler 1990), portfolio

The editor in charge of this paper was Juuso Välimäki.

Acknowledgments: We are grateful to Juuso Välimäki and two anonymous referees for very helpful comments and suggestions throughout the referee process. We also thank Johannes Abeler, Armin Falk, Ori Heffetz, and Matthew Rabin for insightful comments and discussions. The generous support of the National Science Foundation grant # SES-1024683 (Sprenger) is gratefully acknowledged.

E-mail: ugneezy@ucsd.edu (Gneezy); lorenz.goette@uni-bonn.de (Goette); csprenger@ucsd.edu (Sprenger); florian.zimmermann@econ.uzh.ch (Zimmermann)

Journal of the European Economic Association January 2017 15(4):861–876 DOI: 10.1093/jeea/jvw020

choices and the equity-premium puzzle (Benartzi and Thaler 1995; Barberis, Huang, and Thaler 2006), and backward-bending labor supply when wages fluctuate around a daily income target (Camerer et al. 1997; Fehr and Goette 2007).

In all these applications, the reference point was chosen in a plausible, but ad hoc way, with no overarching principle of how to discipline the theory of what reference points are admissible. The reference point was left as a free parameter, implying that tests of reference dependence are truly joint tests of a correct formulation for the reference point and the existence of reference-dependent preferences. Models of expectations-based reference-dependent preferences (Bell 1985; Loomes and Sugden 1986; Gul 1991; Shalev 2000; Kőszegi and Rabin 2006, 2007) eliminate this degree of freedom, easing the incorporation of reference-dependent preferences into broader economic models. Examples of this incorporation include the study of price competition (Kőszegi and Heidhues 2008), contracting in principal-agent settings (Herweg, Müller, and Weinschenk 2010), and daily income targeting (Kőszegi and Rabin 2006; Crawford and Meng 2011). Recent laboratory studies have found support for expectations-based reference dependence (see Abeler et al. 2011; Ericson and Fuster 2011; Gill and Prowse 2012; Banerji and Gupta 2014). This first set of evidence is encouraging, suggesting that expectations play an important role in reference-dependent behavior.

In this paper we build on the design of Abeler et al. (2011), but put the theory to a stricter test. In the experiment of Abeler et al. (2011) subjects work on a real-effort task and earnings are stochastic. The study has two main treatments. In one treatment subjects either earn 7 euros with probability 0.5, or they receive a piece rate. In the other treatment subjects either earn 3 euros, or they receive a piece rate. Uncertainty about earnings is only resolved after subjects have finished working on the task. A key feature of their design is that expectations-based reference dependent incentives change sharply as the outside payments are taken from high to low levels, whereas the neoclassical incentives remain unchanged. In line with the expectations-based reference-dependence model, they find that effort is significantly lower when the outside payment is low.

We move beyond Abeler et al. (2011) along two dimensions. In our experiment, subjects receive a high fixed payment, H, with probability p, they receive a low fixed payment, L, with probability q, and with probability $(1 - p - q) = 0.5$ they are paid according to a piece rate, w. This set-up has two principal motivations. First, for a given set of probabilities, p and q, we can explore the space of outcomes, H and L, beyond the variations in Abeler et al. (2011). Second, for a given set of outcomes, H and L, we can explore the space of probabilities. Similar to variations in fixed payments, the expectations-based reference dependent incentives change sharply as probabilities are changed. Having two possible outside outcomes, H and L, is critical in the design as the neoclassical incentives are unchanged across conditions provided p and q sum to a constant. Note that both the variation in probabilities and the use of high and low fixed payments are novel additions to the Abeler et al. (2011) framework. All our treatments are based on the intuition that the better the outside payment in expectation, the more effort an individual will exert in order to avoid falling short of expectations.

We document two key findings based on the two explorations described above. First, when manipulating the size of fixed payments, expectations-based reference dependence is not sufficient to explain our data. With $p = 0$ and $q = 0.5$, the fixed payment L (H is irrelevant) is varied from \$0 to \$3 to \$7 to \$14. Though, as in Abeler et al. (2011), individuals work less when the fixed payment is \$3 than when it is \$7, when the fixed payment is reduced to \$0 effort actually *increases*. Thus, by varying the fixed payment we do not find a monotone comparative static as predicted by the theory. Second, when manipulating the probabilities on high and low fixed payments, expectations-based reference dependence is also not sufficient to explain our data. In particular, with $L = 0$ and $H = 14$ we find that effort actually *decreases* as the probability, p, of receiving H is increased from 0 to 0.125 to 0.25. By varying the probabilities, we again do not find a monotone comparative static as predicted by the theory.

In sum, our data reveal important deviations from expectations-based reference dependence—as well as standard prospect theory and the neoclassical model. Like Abeler et al. (2011), we replicate the comparative static effects for positive fixed payments. However, our introduction of the zero payment and the comparative statics with respect to the probabilities document clear departures from expectations-based reference points. The findings in Abeler et al. (2011) and ours show clearly that the fixed payments affect effort and therefore affect reference points. However, the general pattern of findings we obtained from our series of treatments shows that these fixed payments affect reference points in more complex ways than predicted by expectations-based reference points.[1] Even though we are unable to offer a full theoretical account for our findings, identifying limitations and falsifying predictions of existing theory is at the heart of scientific progress (Popper 1959).

The remainder of the paper is structured as follows: Section 2 presents the experimental design. In Section 3, we derive theoretical predictions. Section 4 contains results and Section 5 concludes.

2. Design

We build on the experimental design developed in Abeler et al. (2011). The set-up allows a clean manipulation of expectations in a real-effort context. We study a total of nine treatments as summarized in Table 1.

1. In related research, focusing on reference-dependent preferences in exchange experiments, a similar pattern of results can be found: although some studies are favorable of expectations-based reference points (Ericson and Fuster 2011), or at least consistent with it (Smith 2012; Heffetz and List 2014), other papers testing features more closely to the core of expectations-based reference points find clear violations of these predictions (Goette, Harms, and Sprenger 2014).

TABLE 1. Experimental treatments, behavior, and balance.

Treatment (p, q, H, L)	Wage contract	Implementation of uncertainty	Dates	# Subjects	Acquired earnings Mean [median] (s.d.)	Productivity Mean [median] (s.d.)	Time of day Mean [median] (s.d.)	Male (=1) Mean (s.d.)	Asian (=1) Mean (s.d.)
(0, 0.5, NA, 0)	\$0.20 piece-rate (pr = 0.5) or \$0 fixed amount (pr = 0.5)	2 envelopes	Spring 2011/ Winter 2012	30	9.46 [7.8] (6.41)	4.3 [4] (1.68)	13.54 [14.13] (2.5)	.6 (.5)	.8 (.41)
$(0, 0.5, NA, 0)_8$	\$0.20 piece-rate (pr = 0.5) or \$0 fixed amount (pr = 0.5)	8 envelopes	Fall 2011/ Winter 2012	29	8.54 [7] (5.19)	4.17 [5] (1.71)	12.76 [13] (2.48)	.38 (.49)	.79 (.41)
(0, 0.5, NA, 3)	\$0.20 piece-rate (pr = 0.5) or \$3 fixed amount (pr = 0.5)	2 envelopes	Spring 2011/ Winter 2012	30	6.49 [5] (4.23)	4.33 [4] (1.18)	13.5 [14] (2.59)	.57 (.5)	.7 (.47)
(0, 0.5, NA, 7)	\$0.20 piece-rate (pr = 0.5) or \$7 fixed amount (pr = 0.5)	2 envelopes	Spring 2011/ Winter 2012	29	9.08 [7.4] (5.46)	3.76 [4] (1.79)	13.66 [14] (2.57)	.52 (.51)	.76 (.44)
$(0, 0.5, NA, 7)_8$	\$0.20 piece-rate (pr = 0.5) or \$7 fixed amount (pr = 0.5)	8 envelopes	Fall 2011/ Winter 2012	29	8.05 [7] (5.99)	4 [4] (1.6)	12.69 [11.75] (2.54)	.45 (.51)	.79 (.41)
$(0.125, 0.375, 14, 0)_8$	\$0.20 piece-rate (pr = 0.5) or fixed amount of \$0 (pr = 0.375) or fixed amount of \$14 (pr = 0125)	8 envelopes	Fall 2011/ Winter 2012	30	7.35 [5] (5.12)	4.27 [5] (1.84)	13.52 [13.25] (2.68)	.37 (.49)	.7 (.47)
$(0.25, 0.25, 14, 0)_8$	\$0.20 piece-rate (pr = 0.5) or fixed amount of \$0 (pr = 0.25) or fixed amount of \$14 (pr = 0.25)	8 envelopes	Fall 2011/ Winter 2012	29	4.86 [4] (3.85)	4 [4] (1.44)	12.94 [13] (2.62)	.34 (.48)	.69 (.47)
$(0.375, 0.125, 14, 0)_8$	\$0.20 piece-rate (pr = 0.5) or fixed amount of \$0 (pr = 0.125) or fixed amount of \$14 (pr = 0375)	8 envelopes	Fall 2011/ Winter 2012	29	8.34 [9] (5.38)	3.52 [3] (1.66)	12.74 [13] (2.38)	.31 (.47)	.72 (.45)
$(0, 0.5, NA, 14)_8$	\$0.20 piece-rate (pr = 0.5) or \$14 \$14 fixed amount (pr = 0.5)	8 envelopes	Fall 2011/ Winter 2012	30	9.17 [6.2] (7.71)	3.87 [4] (1.66)	12.69 [13] (2.46)	.4 (.5)	.77 (.43)
	H_0: Zero treatment differences. $F(8,256)$				3.02 ($p < 0.01$)	0.89 ($p = 0.52$)	0.83 ($p = 0.58$)	1.28 ($p = 0.25$)	0.30 ($p = 0.97$)

Notes: Summary statistics for nine experimental treatment groups. *Acquired earnings* corresponds to the dollar amount of acquired earnings in Stage 2. *Productivity* corresponds to the number of tables completed in Stage 1. *Time of day* corresponds to the hour (of 24) at which the experiment began. *Male (=1)* and *Asian (=1)* correspond to experimenter's assessment of gender and ethnicity at the time of the experiment. Differences across treatment groups tested with least squares regression of column variable on indicators for treatment assignment with robust standard errors. *F*-statistic for null hypothesis of zero differences across treatments presented in final row.

The experiment consisted of two stages. Subjects received written instructions and had to answer control questions before stage 1 started.[2] Online Appendix C provides the instructions. The complete Online Appendix can be found in the supplementary data. In stage 1 subjects worked for 4 min on a real-effort task. Subjects had to count the number of zeros in a table of 150 zeros and ones. Once a subject had solved a table, a new table of randomly ordered zeros and ones was generated. Subjects were paid ten cents for each table solved correctly. The purpose of stage 1 was to familiarize subjects with the task and to obtain a measure of productivity for each subject.

After subjects finished stage 1, they received instructions for stage 2 and had to answer some control questions. In stage 2 subjects again had to count zeros in tables of zeros and ones. However, in stage 2 subjects could stop working at any time they wanted.[3] To stop working subjects simply had to press a button on the computer screen and inform the experimenter. Once a subject had decided to stop working on the task, he or she received the earnings from the experiment and could leave.

Earnings from stage 2 were determined as follows: with probability p subjects received a high fixed payment, H; with probability q they received a low fixed payment, L; and with probability $(1 - p - q)$ they were paid 20 cents per correctly solved table.[4] Uncertainty about the payment scheme was only resolved after subjects had finished stage 2. Thus, while working on the real effort task, subjects only held expectations regarding their payments. These expectations were manipulated in our nine treatments by varying the parameters p, q, H, and L. All treatments are described by the vector (p, q, H, L). When either $p = 0$ or $q = 0$, the value of either H or L is irrelevant as it is never realized and is described by $H =$ NA or $L =$ NA, respectively.

In our first set of treatments, subjects either received the piece-rate or a fixed payment (each with probability 0.5). In treatments (0, 0.5, NA, 0), (0, 0.5, NA, 3), (0, 0.5, NA, 7), and (0.5, 0, NA, 14), we systematically varied the fixed payment from \$0 to \$14, with intermediate values of \$3 and \$7 (the two values used in Abeler et al. 2011). In all these treatments, uncertainty about payments was introduced as follows: Before stage 2 started, subjects were shown either two or eight envelopes. Half of the envelopes contained a card saying “X dollars”, indicating that subjects

2. Subjects knew in stage 1 that a stage 2 existed but were only provided instructions after stage 1 was complete.

3. Subjects could not work longer than 60 min.

4. Additionally, for every three consecutive incorrect answers submitted, subjects were docked 20 cents. This design element was developed to discourage guessing. In practice, guessing appears to be relatively rare. Out of 265 subjects, 23 (8.7%) ever received a penalty for submitting three incorrect guesses. Of these 23, 19 were penalized once, two were penalized twice, one was penalized three times, and one was penalized four times. In total in the experiment 10,504 tables were correctly solved. Unfortunately, we only record whether a subject was penalized and not their actual number of incorrect guesses. However, assuming each penalty corresponds to three incorrect guesses, this yields a lower bound number of incorrect guesses of $(19 \times 3) + (2 \times 6) + (1 \times 9) + (1 \times 12) = 90$ for these subjects. Assuming all other subjects submit no incorrect guesses this gives a total error rate of $90/(10{,}504+90) = 0.0085$. Though the upper bound can be made arbitrarily high by altering the assumption of behavior for unpenalized observations, the lower bound provided, and the infrequency of being penalized suggests that guessing was indeed relatively rare.

would earn a fixed amount of money (amount depended on treatment), the other half contained a card saying "acquired earnings", indicating that subjects would be paid according to the piece rate. Subjects did not know which card was in which envelope. They were asked to choose one of the envelopes and sign on it. After stage 2 was finished the envelopes were opened and subjects were paid according to the envelope they had chosen.[5]

In three additional treatments, subjects either received the piece rate (with probability 0.5), a low fixed payment of \$0, or a high fixed payment of \$14. In these treatments we varied the probabilities with which subjects received the high or the low payment. In treatments $(0.125, 0.375, 14, 0)_8$, $(0.25, 0.25, 14, 0)_8$, and $(0.375, 0.125, 14, 0)_8$, we gradually increased the likelihood of receiving the high payment. Note that the conditions $(0, 0.5, \text{NA}, 0)_8$ and $(0.5, 0, \text{NA}, 14)_8$ constitute the extremes of this manipulation such that we can trace the probability of \$0 or of \$14 from 0 to 0.5. In all these treatments, uncertainty about the payment was implemented using eight envelopes instead of two to be able to capture the variations in probabilities. For example, in treatment $(0.125, 0.375, 14, 0)_8$ there were four envelopes containing a card saying "acquired earnings", three envelopes containing "0 dollars" and one envelope containing "14 dollars". Again, subjects did not know which card was in which envelope and had to choose one envelope and sign on it before stage 2 started. After stage 2, subjects were paid according to the envelope they had chosen.

We invited only one subject per session for the experiment. Subjects always worked alone on the task and neither saw other participants, nor could they observe or be observed by others. This was done to avoid potential confounds from peer effects (Falk and Ichino 2005) or a preference for conformity (Andreoni and Bernheim 2009). The experiment was conducted at the Rady Behavioral Lab of the University of California, San Diego using the software *z*-tree (see Fischbacher 2007).

In all nine treatments, 29–30 subjects participated, yielding 265 subjects in total. Experimental sessions were run in Spring 2011, Fall 2011, and Winter 2012 with the earliest session on the seventeenth of May 2011 and the latest session on the ninth of March 2012. Within each quarter, subjects were assigned to condition on a rolling basis, in an attempt to ensure balanced collection through time.[6] In Table 1 we note the quarter in which sessions were run. Note that though all nine types of sessions were conducted in Winter 2012, in Spring 2011, and Fall 2011, only a subset of sessions were conducted in each quarter. This may limit comparability across treatments given the potential for differential selection across quarters. In Table 1, we also assess balance on observable characteristics across session types. In addition to actual experimental

5. The experiments were conducted with both two envelopes (as in the original experiments) and eight envelopes (to ensure comparability with the conditions where probability weights were varied). Table 1 indicates treatments where eight envelopes were used by $(p, q, H, L)_8$. As no qualitative differences are obtained across subtreatment for envelope number, the analysis combines the data across two and eight envelope subtreatments and controls for the implementation of uncertainty.

6. This balance appears to have been achieved. Controlling for the quarter, we find no systematic differences in the date of study across the nine treatment groups, ($F(8, 254) = 0.88$, $p = 0.53$), in a regression of date on indicators for treatment and indicators for quarter with robust standard errors.

behavior, we present *Productivity*, calculated as the number of tables completed during Stage 1; *Time of Day*, the hour at which the session began (out of 24); and *Male (=1)* and *Asian (=1)*, based upon the experimenter's assessment of gender and ethnicity at the time of the study. Though some qualitative differences in gender are observed across sessions, for none of the observables can we reject the null hypothesis of equal assignment across treatments. This lack of difference in observables is contrasted with the sharp and significant differences in *Acquired Earnings* across treatments. In all analysis we control for the observable characteristics noted in Table 1 along with the quarter in which the session was conducted and the implementation of uncertainty.

3. Theoretical Predictions

We consider optimal effort provision within our experimental framework for individuals with standard preferences and individuals with expectations-based reference-dependent preferences. Wage contracts in our experiment take the following shape: with probability p subjects are paid a high fixed amount, H, regardless of effort; with probability q subjects are paid a low fixed amount, $L < H$, regardless of effort; and with probability $1 - p - q$ subjects are paid a piece rate, w, for every unit of exerted effort, e.

Individuals set effort, e^*, to equate marginal costs with expected marginal benefits. The key insight from expectations-based reference dependence is that variations in p, q, H, and L induce kinks in the marginal benefit of effort even when the key neoclassical variable of interest, $(1 - p - q) \cdot w$, remains fixed. Hence, although standard models predict the derivatives $\partial e^*/\partial L$ and $(\partial e^*/\partial p)_{1-p-q}$ to be zero, expectations-based reference dependence delivers signed predictions on these comparative statics.[7]

To fix ideas in the standard model, we assume an additively separable linear-in-earnings utility function,[8] where the total utility of an effort choice under a given wage contract is

$$U(e) = (1 - p - q) \cdot we + pH + qL - c(e),$$

where $c(e)$ is an assumed strictly convex cost of effort function, $c'(e) > 0$, $c''(e) > 0$. Optimal effort is set to equate marginal costs and benefits in the standard model,

$$MB_S(p, q) = (1 - p - q)w = c'(e) = MC,$$

7. The subscript of $(\partial e^*/\partial p)_{1-p-q}$ refers to keeping fixed the value $(1 - p - q)$ such that q is changing by an equal and opposite amount to p. We maintain the partial derivative notation for both $(\partial e^*/\partial p)_{1-p-q}$ and $\partial e^*/\partial L$ when the solution function is nondifferentiable only for convenience. See Online Appendix A.1 for detail.

8. We assume linearity in earnings, we, for ease of exposition. The predictions would be similar if we assumed some curvature or even loss aversion around a status-quo reference point. The linearity could also be thought of as a small stakes assumption under expected utility.

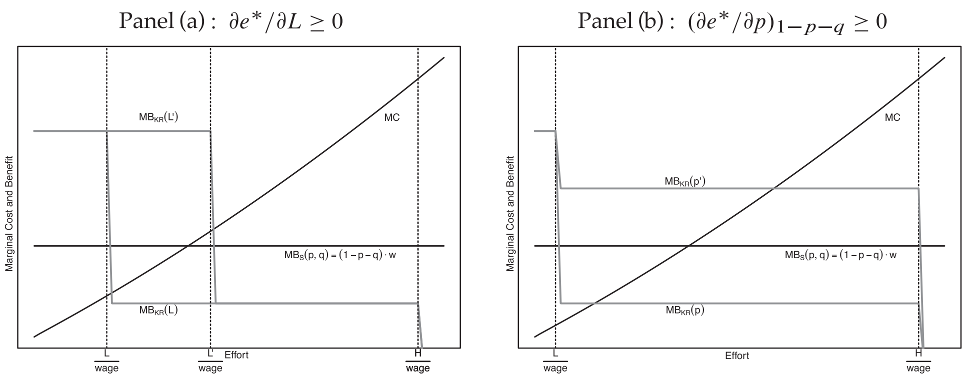

FIGURE 1. Comparative statics.

as in the intersection of the curves $MB_S(p, q)$ and MC in Figure 1. Note that for constant $1 - p - q$, regardless of the values of p, q, H, or L, the marginal benefits of effort are fixed. Hence, $\partial e^*/\partial L = 0$ and $(\partial e^*/\partial p)_{1-p-q} = 0$ in the standard model. The intuition is straightforward: since L is paid only in the state where effort does not matter, naturally, effort should not be affected by it. By the same reasoning, varying p has no impact, since it only affects the payoff in case the effort choice is irrelevant.

Similarly, if individuals have reference-dependent preferences, but their reference point is given by, for example, an earnings target for the experiment and does not depend on the expectations created within our experiment, manipulating L or p should also have no effect on behavior—the logic is the same as above.

In sharp contrast, expectations-based reference dependence requires that changes in p, q, H, and L alter behavior as they alter expectations and the perceived marginal benefits of effort. To describe the effects we use the model of Kőszegi and Rabin (2006, 2007) (henceforth KR), which features expectations-based reference dependence around a stochastic reference distribution. We begin by analyzing the predictions under KR preferences, making the commonly used assumption of a linear gain–loss function, and employing the equilibrium concept, "Choice Acclimating Personal Equilibrium". Online Appendix A provides full mathematical detail for this formulation and provides several extensions to ensure robustness of our theoretical predictions. These extensions are discussed at the end of this section.

Under KR preferences, individuals assess the probability that each outcome will exceed or fall short of each potential expectation. In our environment an individual expects H with probability p, L with probability q, and we with probability $1-p-q$. Total utility is the expected utility over these combinations of expectations and outcomes captured by

$$U(e) = p \cdot E[u(e|H)] + q \cdot E[u(e|L)] + (1 - p - q) \cdot E[u(e|we)] - c(e),$$

where $E[\cdot]$ is the expectation operator taken over possible outcomes for a given referent.[9] For example,

$$E[u(e|H)] = pu(H|H) + qu(L|H) + (1 - p - q)u(we|H).$$

Following KR, we define the reference-dependent utility $u(x|r) = m(x) + \mu(m(x) - m(r))$, where $m(\cdot)$ is assumed to be linear and a piecewise linear gain–loss utility function is adopted,

$$\mu(y) = \begin{cases} \eta \cdot y & \text{if } y \geq 0, \\ \eta \cdot \lambda \cdot y & \text{if } y < 0. \end{cases}$$

In this formulation η represents gain utility above the referent and $\eta\lambda$ represents loss utility below the referent, with $\lambda > 1$ indicating loss aversion.

From the formulation above, it is clear that the marginal returns to effort will depend on the relationship between accumulated earnings, we, and H, and L, and the probabilities that each are realized. We can see the intuition as inducing kinks in the marginal benefits of effort. For $we < L$, increasing effort reduces the severity of loss relative to both L and H, encouraging effort. The returns to effort are diminished once $we = L$, as increasing effort reduces the severity of losses associated with expecting H and receiving we, but increases the severity of losses of expecting we and receiving L. The returns to effort are diminished again at $we = H$, as increasing effort increases the severity of losses associated with expecting we and receiving either L or H. Hence, the marginal benefit function can be seen as a step function with pronounced kinks at $e = L/w$ and $e = H/w$ related to the degree of loss aversion and the probability of losses. Figure 1 provides graphical detail and Online Appendix A.1 provides the algebra.

From this basis we can construct the comparative statics of interest. First, given the kinks in marginal benefit at L/w and H/w changes to these values will have an influence on observed behavior. We sign $\partial e^*/\partial L \geq 0$ for a convex cost function. For a given p, q, and H, effort weakly increases in the low payment, L. Figure 1, Panel (a) provides the graphical detail and Online Appendix A.1 provides the algebra.[10]

Second, changes in the probabilities p and q for a fixed $1 - p - q$ change the severity of the kinks in marginal benefit. In the extremes, at $p = 0$ there is no kink at H/w, individuals only have negative marginal incentives above $we = L$. Further for $q = 0$ there is no kink at L/w and individuals have only positive marginal incentives until $we = H$. We sign $(\partial e^*/\partial p)_{1-p-q} \geq 0$ for a convex cost function.

9. Note that decision makers treat expectations and outcomes independently such that for each possible expectation, the decision maker considers all possible outcomes. In motivating this structure, Kőszegi and Rabin (2007) cite a literature in psychology indicating that individuals engage in counterfactual thinking when considering unwon prizes.

10. Figure 1, Panel (a) demonstrates that those individuals whose marginal cost intersects the kink in $MB(L)$ at L/w will have $MB(L') > MC$ at L/w when the low payment increases to L'.

The higher the value of p, holding $1 - p - q$ constant, the more likely it is that we should expect H, the higher are the marginal incentives to work below $we = H$. Figure 1, Panel (b) provides the graphical detail and Online Appendix A.1 provides the algebra.[11]

The two comparative statics lead to our two core predictions.

PREDICTION 1. *For a fixed H, p, and q, effort is increasing in L such that effort should be monotonically increasing in (0, 0.5, NA, 0), (0, 0.5, NA, 3), (0, 0.5, NA, 7), (0, 0.5, NA, 14).*

PREDICTION 2. *For a fixed $1 - p - q$, L, and H, effort is increasing in the probability of the high outcome such that effort should be monotonically increasing in (0, 0.5, NA, 0), (0.125, 0.375, 14, 0), (0.25, 0.25, 14, 0), (0.375, 0.125, 14, 0), (0, 0.5, NA, 14).*

Our design is focused on testing these two theory-driven predictions. To ensure that these predictions are robust, Online Appendix A provides a variety of extensions. In Online Appendix A.2, we relax the assumption of linearity in gain–loss utility, allowing for "diminishing sensitivity" to gains and losses. The predictions of increasing effort in response to increases in L and p are maintained, provided that diminishing sensitivity is more extreme for gains than for losses.[12] In Online Appendices A.3 and A.4, we analyze two alternative formulations for equilibrium effort level choices. First, we consider an alternative equilibrium concept, termed "Personal Equilibrium" (PE), wherein a specific level of effort can be supported as a PE choice provided the decision maker would choose such a level of effort over all others if she expected to do so. The potential multiplicity of PE effort levels requires analysis of the set of such choices. We show that the sets of supportable PE choices are influenced by L and p in intuitive ways. Higher effort levels grow easier to support as PE choices and lower effort levels grow more difficult to support as L and p increase. Second, we recognize the inherent stopping problem nature of our effort decisions, and examine whether a given effort level can be supported relative only to higher effort levels. A similar logic maintains: increasing L and p increases the lowest effort level at which the decision maker can support stopping. In Online Appendix A.5, we provide an alternative formulation for expectations-based reference dependent effort choice based on the disappointment aversion models of Bell (1985) and Loomes and Sugden (1986). Our two predictions are again maintained. Together, these exercises help to ensure the robustness of our theoretical developments to alternate functional forms and alternate formulations for how effort levels are chosen.

11. Figure 1, Panel (b) demonstrates that the marginal cost, MC, intersects the marginal benefit curve $MB_{KR}(p')$ at a higher value of effort than for $MB_{KR}(p)$, where $p' > p$ and $1 - p' - q' = 1 - p - q$.

12. The condition of more extreme diminishing sensitivity for gains relative to losses is a sufficient condition and can be relaxed for empirically relevant cases. See Online Appendix A.2 for further detail.

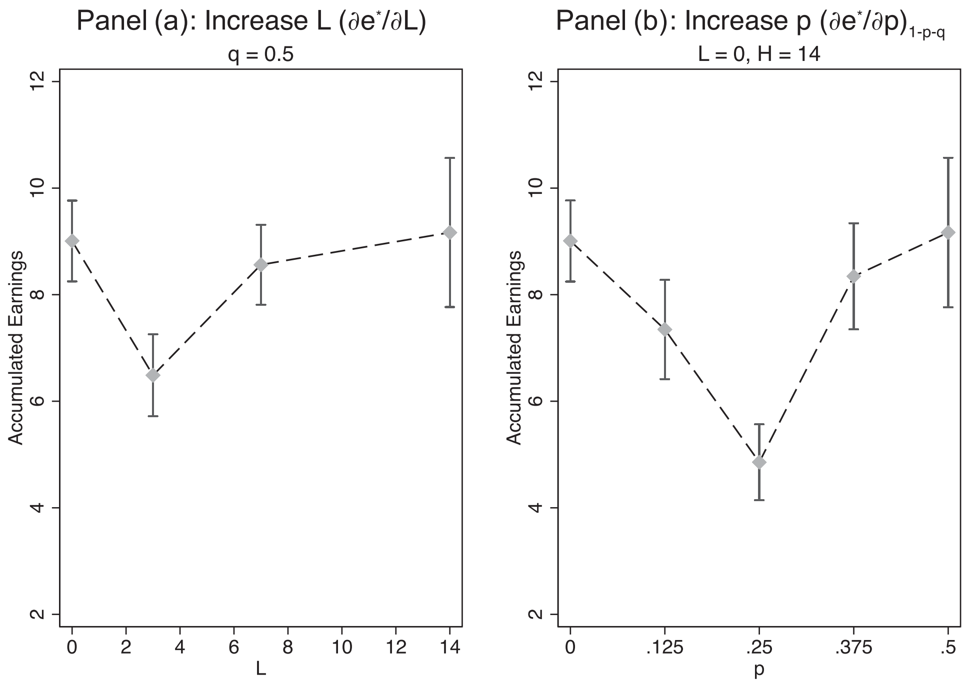

FIGURE 2. Average accumulated earnings across treatments. Standard error bars corresponding to +/− one robust standard error. Panel (a): Average accumulated earnings for each value of L from treatments (0, 0.5, NA, L). Panel (b): Average accumulated earnings for each value of p from treatments (p, q, 14, 0). Observations from subtreatments (0, 0.5, NA, 0) and $(0, 0.5, \text{NA}, 0)_8$ as well as for subtreatments (0, 0.5, NA, 7) and $(0, 0.5, \text{NA}, 7)_8$ are combined.

4. Results

A total of 265 subjects participated in our nine conditions. Over all treatments, subjects on average solved 39.64 tables, leading to accumulated earnings of $7.93. The average time subjects worked on the task was 33.30 min. Table 1 provides means and standard deviations for all experimental conditions and Figure 2 summarizes our key findings. The key measure of effort will be *Accumulated Earnings*, we^*. Accumulated earnings are graphed against the expected value of the manipulated payments, $pH + qL$, with separate series for our two predictions.

4.1. Changing Outcomes

To test Prediction 1, we first consider treatments (0, 0.5, NA, 3) and (0, 0.5, NA, 7), the two treatments from Abeler et al. (2011). Figure 2, Panel (a) and Table 1 show that, as predicted by theories of expectations-based reference dependence and found in Abeler et al. (2011), effort increases as we move from a fixed payment of $3 to a fixed payment of $7. In the two envelope conditions, subjects on average stop working at accumulated earnings of \$9.08 when the fixed payment is \$7. When the fixed payment is \$3, the average stopping point is \$2.59 lower. This difference is significant using a Ranksum-test ($z = 2.24$, $p < 0.05$). Beyond the fixed payments of \$3 or \$7, Figure 2, Panel (a) traces out the response of effort to changes in the value of L. Interestingly, the series reveals a U-shape relationship between L and exerted effort. When the fixed outside payment is reduced to zero (treatment (0, 0.5, NA, 0)), effort is *higher* than when the outside payment is \$3 in the two envelope condition (difference = \$2.97, $z = 1.90$, $p < 0.10$).[13]

Table 2 analyzes our comparative statics in a regression framework. Accumulated earnings are regressed on indicators for treatments separated by our two comparative statics of interest. In column (1), Accumulated Earnings are regressed on indicators for the four treatments delivering the comparative static $\partial e^*/\partial L$. Additional controls for observable characteristics, the quarter of study and the implementation of uncertainty are provided in column (2). Table 2, columns (1) and (2) reveal an interesting nonmonotonicity to behavior. From a baseline of around \$9 in accumulated earnings when the fixed payment is \$0, effort decreases substantially when the fixed payment is \$3 and then rebounds at higher levels. Although we marginally reject the null hypothesis of zero treatment differences predicted by standard models, our results suggest a nonmonotonicity not captured by expectations-based reference dependence.

For changes in L we both reject the standard model and identify limitations for expectations-based reference dependence. Effort decreases from (0,0.5, NA, 0) to (0, 0.5, NA, 3) and increases from (0, 0.5, NA, 3) to (0, 0.5, NA,14).

RESULT 1. *We reject the prediction of a monotone increase in effort in response to an increase in L.*

As models of expectations-based reference dependence predict a monotone response to the outside payment L, the observed nonmonotonicity represents a first contradiction of the model. Interestingly, this violation is obtained at the limit $L = 0$, where expectations-based reference dependence generates only negative gain-loss incentives for work. That no such negative effect on effort is observed is suggestive evidence that the low outside payment is somewhat ignored when setting effort. Part of the discussion with respect to the evidence of Abeler et al. (2011) was the possibility that subjects make value inferences when told the amount of the outside payment (Ericson and Fuster 2011). A \$3 fixed payment indicates an unimportant task generating lower effort than a \$7 fixed payment. A natural interpretation of such a signaling argument would imply that the value inference from a \$0 fixed payment is even lower, which our data seems to reject.

13. Online Appendix Figure A.1, panel A presents histograms of behavior in each relevant condition. Following the mean data, the distribution shifts toward decreased *Accumulated Earnings* as the value of L is increased from \$0 to \$3.

TABLE 2. Comparative static regressions.

	(1)	(2)	(3)	(4)
	$\partial e^*/\partial L$		$(\partial e^*/\partial p)_{1-p-q}$	
	Dependent variable: Accumulated earnings			
(0, 0.5, NA, 3)	−2.520**	−3.193**		
	(1.080)	(1.345)		
(0, 0.5, NA, 7)	−0.445	−0.091		
	(1.068)	(1.056)		
(0.125, 0.375, 14, 0)			−1.660	−1.517
			(1.203)	(1.294)
(0.25, 0.25, 14, 0)			−4.152***	−3.730***
			(1.042)	(1.149)
(0.375, 0.125, 14, 0)			−0.662	0.369
			(1.253)	(1.283)
(0, 0.5, NA, 14)	0.160	1.039	0.160	0.961
	(1.592)	(1.495)	(1.596)	(1.508)
Constant: (0, 0.5, NA, 0)	9.007***	8.358***	9.007***	6.674**
	(0.759)	(3.029)	(0.761)	(2.883)
Implementation (8 envelopes)		0.692		0.841
		(1.854)		(2.830)
Productivity		1.060***		1.174***
		(0.336)		(0.280)
Time of day		−0.156		0.032
		(0.177)		(0.169)
Male (=1)		−0.381		−1.155
		(0.882)		(0.950)
Asian (=1)		−1.022		−1.853*
		(1.139)		(0.994)
Fall 2011 (=1)		−2.577		−2.271
		(2.037)		(3.081)
Winter 2012 (=1)		−1.012		−2.620
		(1.492)		(2.960)
No. observations	177	177	177	177
R-squared	0.024	0.125	0.067	0.200
H_0: (0, 0.5, NA, 3) = (0, 0.5, NA, 7)	$F(1, 173) = 3.73$	$F(1, 166) = 5.43$		
Abeler et al. (2011)	($p < 0.10$)	($p < 0.05$)		
H_0: zero treatment difference	$F(3, 173) = 2.26$	$F(3, 166) = 2.32$	$F(4, 172) = 4.92$	$F(4, 165) = 4.53$
Neoclassical prediction	($p < 0.10$)	($p < 0.10$)	($p < 0.01$)	($p < 0.01$)

Notes: Coefficients from ordinary least squares regression. Robust standard errors are given in parentheses. Additional controls in columns (2) and (4). *Implementation* corresponds to the implementation of uncertainty. *Productivity* corresponds to the number of tables completed in Stage 1. *Time of day* corresponds to the hour (of 24) at which the experiment began. *Male (=1)* and *Asian (=1)* correspond to experimenter's assessment of gender and ethnicity at the time of the experiment. *Fall 2011 (=1)* and *Winter 2012 (=1)* are indicators for the quarter in which the session took place.

Level of significance: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

4.2. Changing Probabilities

To test Prediction 2 we first examine the series from Figure 2, Panel (b) that captures the increase in p. Expectations-based reference dependence predicts a monotone effect of our variations in probabilities on effort. Again we obtain a U-shaped relationship between expectations and effort. As the probability of receiving a high outcome increases, holding fixed neoclassical incentives, effort decreases, reaches a minimum in condition (0.25, 0.25, 14, 0), and then rebounds. Table 2 columns (3) and (4) provide corresponding regression analyses. From a baseline of around \$9 in accumulated earnings for condition (0, 0.5, NA, 0), effort is both substantially and significantly reduced as p increases to 0.25 in condition (0.25, 0.25, 14, 0), and then effort increases thereafter. Across specifications we strongly reject the neoclassical null hypothesis of zero condition differences. Subjects work around half as much when the outside payment is a 50–50 mixture over \$0 and \$14 than when the outside payment is \$0 for sure or \$14 for sure.[14]

For changes in p, we both reject the standard model and identify limitations for expectations-based reference dependence. Effort decreases from (0, 0.5, NA, 0) to $(0.25, 0.25, 14, 0)_8$ and increases from $(0.25, 0.25, 14, 0)_8$ to (0, 0.5, NA, 14). The observed nonmonotonicity to behavior is striking and in contrast to both neoclassical predictions and predictions of expectations-based reference dependence.

RESULT 2. *We do not find support for a monotone increase in effort in response to an increase in the probability of a high payment, p.*

5. Concluding Remarks

We test the limits of expectations-based reference-dependent preferences (Bell 1985; Loomes and Sugden 1986; Gul 1991; Shalev 2000; Kőszegi and Rabin 2006, 2007). We build on the design by Abeler et al. (2011), that has been used to test (and found support for) models of expectations-based reference-dependence. Our set-up allows clean manipulation of expectations, and simple model-driven comparative statics for the effects on effort provision of increasing fixed outside payments and changing relative probabilities of outside payments. Our manipulations hold fixed the key neoclassical object of interest, the expected piece-rate wage.

As a whole, our findings suggest that individuals display reference-dependent preferences, and that their reference points are affected by possible fixed payments. Current models formulate the force of such fixed payment through expectations-based reference points. However, our results show such expectations-based formulations are insufficient for understanding all of reference-dependent behavior.

Though the final account will necessarily be hashed out in future work, our findings have implications for which theoretical developments will prove promising. One natural avenue for potential future research revolves around ways in which sensations of gains and losses may not only be dependent on expectations but also on other contextual variables, such as the available set of actions. First, when we provide individuals with external payments outside of what they will achieve, such as 0 dollars, these values appear to be ignored. A plausible constraint on the formation of reference points could

14. Online Appendix Figure A.1, Panel A presents histograms of behavior in each relevant condition. Following the mean data, the distribution shifts towardsdecreased *Accumulated Earnings* as p is increased from 0 to 0.25 and then shifts toward increased *Accumulated Earnings* thereafter.

be that the referent must be an outcome that is not immediately surpassed for it to be attended to. Second, when we provide referents with residual uncertainty, such as with (0.25, 0.25, 14, 0), effort reduces relative to conditions in which certainty in payments can be attained by tuning effort, such as (0, 0.5, NA, 7). Perhaps if certainty cannot be attained with the actions available to the decision maker it is particularly discouraging.[15] These two observations carry a common theme: the set of actions and corresponding distribution of outcomes may influence both what is held as a referent and the extent to which losses are felt when falling short thereof. Future research linking reference dependence to available actions may well prove fruitful both for theoretical development and for empirical analyses experimentally altering choice sets and examining corresponding effects for the extent of reference dependent behavior.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman (2011). "Reference Points and Effort Provision". *American Economic Review*, 101(2), 1–25.

Andreoni, James and Douglas B. Bernheim (2009). "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica*, 77, 1607–1636.

Banerji, A. and Neha Gupta (2014). "Detection, Identification, and Estimation of Loss Aversion: Evidence from an Auction Experiment." *American Economic Journal: Microeconomics*, 6, 91–133.

Barberis, Nicholas, Ming Huang, and Richard H. Thaler (2006). "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing." *American Economic Review*, 96(4), pp. 1069–1090.

Bell, David E. (1985). "Disappointment in Decision Making under Uncertainty." *Operations Research*, 33, 1–27.

Benartzi, Shlomo and Richard H. Thaler (1995). "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics*, 110, 73–92.

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

Crawford, Vincent and 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.

Ericson, Keith and Andreas Fuster (2011). "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments." *Quarterly Journal of Economics*, 126, 1879–1907.

Falk, Armin and Andrea Ichino (2005). "Clean Evidence on Peer Effects." *Journal of Labor Economics*, 34, 39–57.

Fehr, Ernst and Lorenz Goette (2007). "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment". *American Economic Review*, 97(1), 298–317.

Fischbacher, Urs (2007). "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10, 171–178.

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

15. This resonates with the proposed rationale for anomalous risk-taking behaviors like the Gneezy, List, and Wu (2006) "uncertainty effect" where a 50–50 gamble over book-store gift certificates is valued less than the certainty of the gamble's worst outcome. Gneezy et al. (2006) indicate that uncertainty per se may carry with it a utility penalty.

Gill, David and Victoria Prowse (2012). "A Structural Analysis of Disappointment Aversion in a Real Effort Competition". *American Economic Review*, 102(1), 469–503.

Gneezy, Uri, John A. List, and George Wu (2006). "The Uncertainty Effect: When a Risky Prospect Is Valued Less Than Its Worst Possible Outcome." *The Quarterly Journal of Economics*, 121, 1283–1309.

Goette, Lorenz, Annette Harms, and Charles Sprenger (2014). "Randomizing Endowments: An Experimental Study of Rational Expectations and Reference-Dependent Preferences." *Working Paper*, University of Bonn.

Gul, Faruk (1991). "A Theory of Disappointment Aversion." *Econometrica*, 59, 667–686.

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

Herweg, Fabian, Daniel Müller, and Philipp Weinschenk (2010). "Binary Payment Schemes: Moral Hazard and Loss aversion". *American Economic Review*, 100(5), 2451–2477.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk". *Econometrica*, 47(2), 263–291.

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

Kőszegi, Botond and Matthew Rabin (2007). "Reference-Dependent Risk Attitudes". *American Economic Review*, 97(4), 1047–1073.

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

Kőszegi, Botond and Paul Heidhues (2008). "Competition and Price Variation when Consumers are Loss Averse". *American Economic Review*, 98(4), 1245–1268.

Loomes, Graham and Robert Sugden (1986). "Disappointment and Dynamic Consistency in Choice under Uncertainty." *The Review of Economic Studies*, 53, 271.

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

Popper, Karl (1959). *The Logic of Scientific Discovery*. Basic Books, New York.

Shalev, Jonathan (2000). "Loss Aversion Equilibrium." *International Journal of Game Theory*, 29, 269–287.

Shefrin, Hersh and Meir Statman (1985). "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence." *The Journal of Finance*, 40, 777–790.

Smith, Alec (2012). "Lagged Beliefs and Reference-Dependent Preferences." *Working Paper*, Caltech.

SUPPLEMENTARY DATA

Supplementary data are available at *JEEA* online.