Behavioural Economics **University of Oxford**
Vincent P. Crawford **Michaelmas Term 2012**

Behavioural Decision Theory: Mugs, Cabs, and Reference-Dependence

(based in part on Vincent Crawford and Juanjuan Meng; "New York City Cabdrivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," 2011 *AER*; and with large debts to her)

Revised 9 August 2012

Mugs and the willingness to pay-willingness to accept gap

In a famous experiment, Kahneman, Knetsch, and Thaler ("KKT"; 1990 *Journal of Political Economy*) randomly gave mugs to half the subjects in a classroom experiment ("owners") and nothing to the others ("nonowners").

They then elicited selling and buying prices for owners or nonowners, using a procedure that gives subjects an incentive to reveal their true prices:

- Subjects are told that a price has been selected randomly, and is sealed in an envelope in front of the room (in plain view of all).
- They then get a sheet of paper with a bunch of possible prices listed, and they are asked to indicate whether they would buy at each price.
- The highest price at which a buyer expresses a willingness to buy is taken as her/his "buying price".
- The highest price at which a seller expresses a willingness to keep the good is taken as her/his "selling price".

In the field there might have been selection effects:

Mug owners might have had higher prices than nonowners on average just because they were the ones who chose to acquire them, or perhaps because they had learned to love their mugs.

Owners might also have known more about mug quality than nonowners.

But in the experiment owners and nonowners were randomly assigned, and all had equal opportunity to inspect the mugs.

Thus in a large enough sample, with a common value distribution for owners and nonowners, supply and demand should be mirror images of each other.

However, the average buying price of nonowners was about $3.50, and the average selling price of owners was about $7.00:

Way too big a willingness to pay-willingness to accept gap to be random.

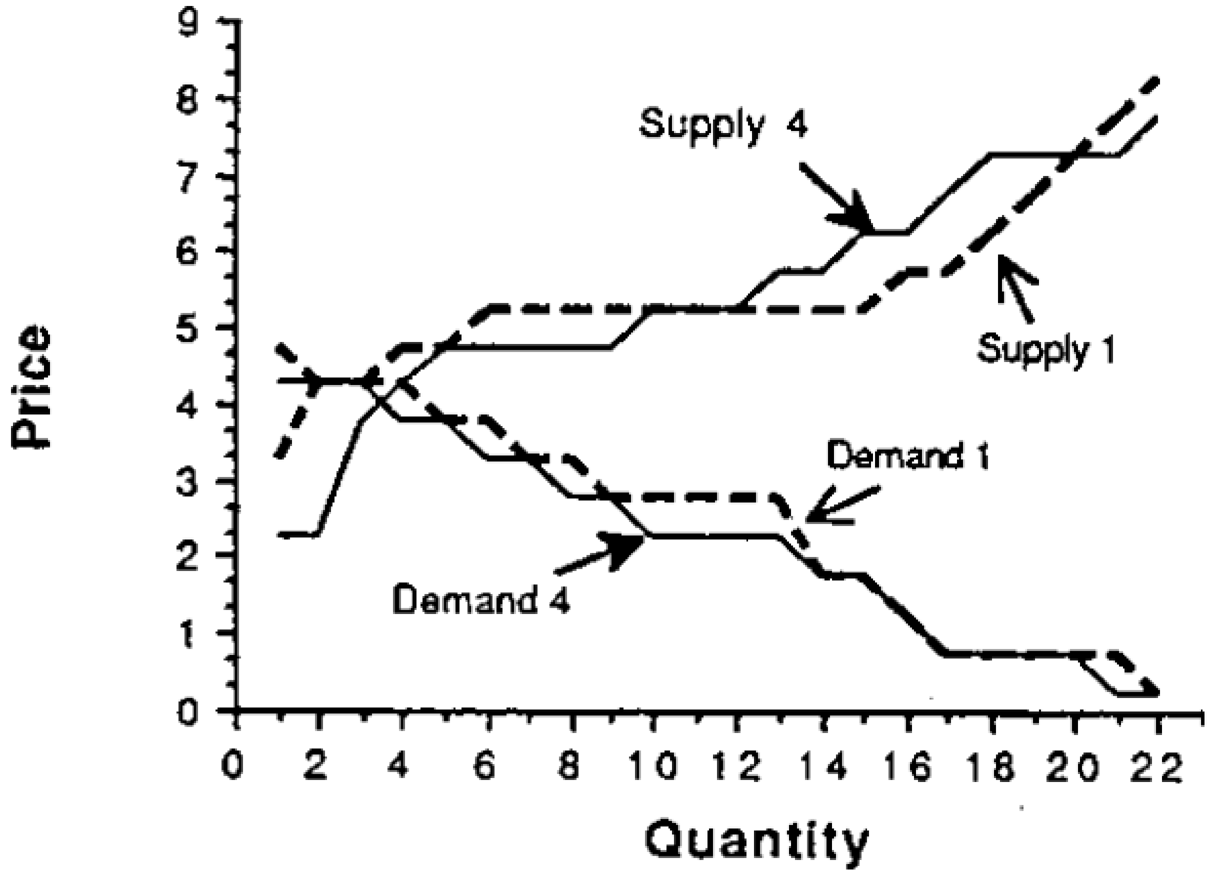

FIG. 1.—Supply and demand curves, markets 1 and 4

This result has been replicated many times, with the gap almost always large and in the same direction.

(But see the "maybe-no-gap" findings of Plott and Zeiler, 2005, 2007 *AER* and the 2011 *AER* Comment and Reply by Isoni et al. and Plott and Zeiler.)

Standard neoclassical demand theory would model this situation by postulating a utility function over *levels* of consumption of mugs and/or money (treating money as a proxy for other lifetime consumption).

In such a model a person’s willingness to pay money for the mug should approximately equal his willingness to accept money for it.

But subjects who received mugs were on average slightly richer than those who did not, and in theory even a tiny change in income or wealth can radically alter a person’s mug-money tradeoff.

Thus in a neoclassical model of choice among levels it is *logically* possible that the gap is due to income effects.

But there’s no reason why people with varying initial wealths should all (or even most) have income effects that on average double their values for mugs, right at the person’s status quo income level before the experiment.

It would violate econometric modeling conventions to allow such a magical correlation between initial wealth and how mug value relates to income.

Further, even if we allowed such a correlation, income effects from mugs are not large enough to plausibly explain such a large gap.

Thus it seems that no reasonable specification of preferences over *levels* of mug and money consumption *alone* will make income effects in a neoclassical model a credible explanation of the gap.

Aside:

Income effects can be ruled out more definitively as an explanation by an experiment in which another group of subjects, “choosers,” are told they will be given either a mug or money, and asked to state the amount of money that makes them indifferent between the amount and the mug.

Choosers have the same incentives to reveal their true “reservation price” for the mug that sellers in the original KKT experiment did.

Yet in a typical experiment, the average selling, buying, and choosing prices were $7.12, $2.87, and $3.12 respectively.

Thus choosers, who have approximately the same income as owner/sellers (because they know they are going to get either a mug or at least an equivalent amount of money), still have reservation prices approximately like buyer/nonowners, who have no such income.

End of aside

Cabs

Camerer, Babcock, Loewenstein, and Thaler (1997 *QJE*) studied the labor supply of New York City cabdrivers.

Cabdrivers are ideal for testing theories of labor supply because, unlike most workers in modern economies, they choose their own hours each day, and conditions are roughly constant within a day.

Camerer et al. found a strongly negative elasticity of daily hours with respect to realized daily earnings.

Realized daily earnings (or realized expected daily earnings) is the natural counterpart of the wage in this setting, and is henceforth called the “wage”.

The correlations between log hours and log wages (also known as elasticities) in various groups were between -0.503 and -0.269:

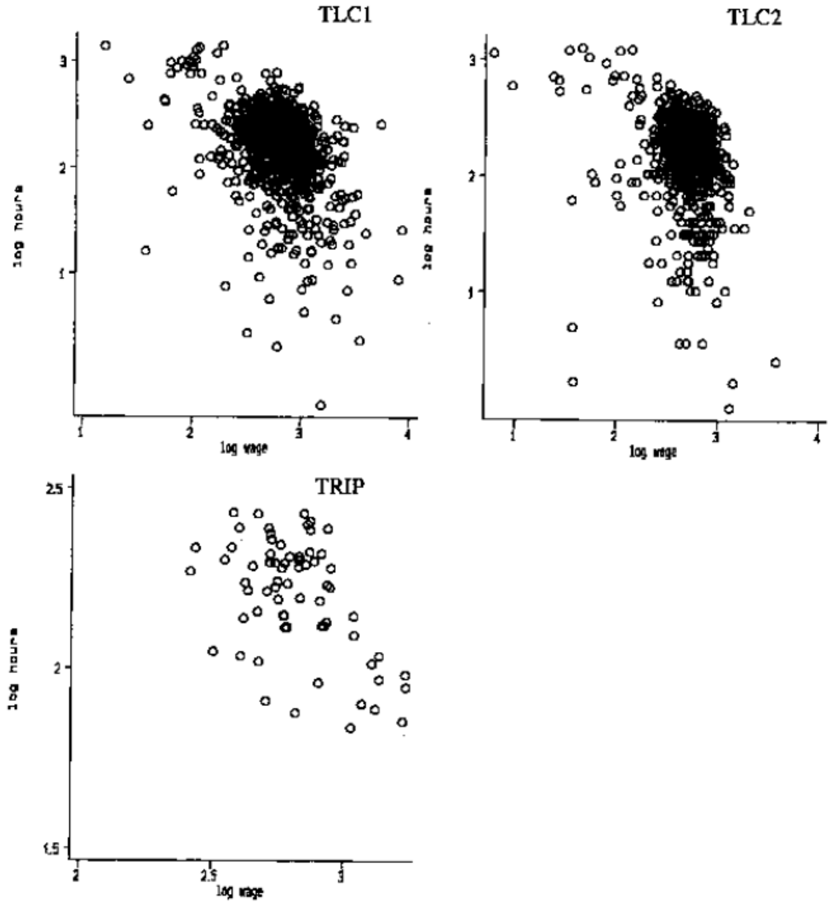

FIGURE I
Hours-Wage Relationships

Thus drivers work less on days when the wage is high.

This negative elasticity reduces earnings: If you reach the target very early, it strongly suggests (treating realized earnings as an indicator of earnings later that day) that you could easily earn a lot more by working longer.

As a result neoclassical choice theory predicts a positive wage elasticity of hours, because the income effects of changes in the wage are negligible.

To explain their results Camerer et al. informally proposed a model in which drivers have daily income targets and work until the target is reached.

Drivers therefore tend to work less on days when the wage is high.

(Literal income targeting would yield elasticity -1; noise pushes it toward 0.)

Inspired by Camerer et al.'s study, Farber (2005 *JPE*) collected and analyzed data on a new set of New York City cabdrivers, finding that:

- Before controlling for driver fixed effects, the probability of stopping work is significantly related to income realized on a given day, but
- Driver fixed effects and other relevant controls render this effect statistically insignificant, and
- The probability of stopping is significantly related to cumulative hours.

Farber (2008 AER) used his 2005 dataset to estimate a structural model with explicit daily income targets.

He estimated drivers' (not directly observable) targets as latent variables with driver-specific means and driver-independent variance.

He assumed, mainly for computational reasons, that both mean and variance of income are constant across days of the week, thus allowing the target to vary across days for a given driver, but only as a random effect.

This assumption is strongly rejected in the data, with Thursdays' through and Sundays' mean incomes systematically higher than those of other days.

Farber included day-of-the-week dummies in his main specifications of the stopping probability equation, but this turns out to be an imperfect substitute for allowing the income target to vary across days of the week.

Farber (2008 AER) found that a sufficiently rich parameterization of his income-targeting model has a better fit than a standard neoclassical model.

The estimated probability of stopping increases significantly and substantially once the income target is reached.

But the estimated model cannot reconcile the strong increase in stopping probability at the target with the aggregate smoothness of the relationship between stopping probability and realized income.

Further, the random effects in drivers' targets have very high estimated variances, from which he concluded that income targets are too unstable and imprecisely estimated to yield a useful reference-dependent model.

Finally, Farber’s model with estimated latent targets does not clearly distinguish anticipated, permanent from transitory wage increases.

Such a distinction is an important part of the story, because it’s hard to believe that an anticipated increase in the wage would reduce labor supply.

And other studies of workers who choose their own hours have found positive relationships between *expected* earnings and labor supply.

- Oettinger (1999 *JPE*) finds that stadium vendors are more likely to go to work on days when their wage can be expected to be higher; and

- Fehr and Goette (2007 *AER*) find that bicycle messengers sign up for more shifts when their commissions are experimentally increased.

As Farber (2008 *AER*) suggests, a finding that labor supply is reference-dependent would have significant policy implications:

> "Evaluation of much government policy regarding tax and transfer programs depends on having reliable estimates of the sensitivity of labor supply to wage rates and income levels. To the extent that individuals' levels of labor supply are the result of optimization with reference-dependent preferences, the usual estimates of wage and income elasticities are likely to be misleading."

Even if the effects of reference-dependence wash out on average, ignoring it when it is really there (as has been the custom in labor economics) may yield biased estimates even of the models' "neoclassical" coefficients.

But despite a number of empirical studies, the literature has not converged on the extent to which the evidence supports reference-dependence.

Prospect theory and reference-dependent preferences

The willingness-to-pay-willingness-to-accept gap and cabdrivers' negative wage elasticity of labor supply, which are not obviously related, can both be gracefully explained by allowing preferences to be reference-dependent.

Reference-dependence expands the space over which preferences are defined to include not only levels but also *changes* in consumption, which must be measured as gains or losses relative to some reference point.

Because the vN-M axioms are silent on what preferences are about, the assumption that preferences respond to changes as well as levels is no less consistent with expected-utility maximization than the assumption that only levels matter.

Kahneman and Tversky (1979 *Econometrica*) focused on preference over changes alone, as may be appropriate for small-stakes lab experiments.

But I will follow Kőszegi and Rabin's (2006 *QJE*) implementation of prospect theory, in which preferences respond to both changes and levels.

Aside:

The most plausible escape routes other than reference-dependence are closed off:

- Ambiguity aversion doesn't help, because the reactions that cause the problem are reactions to known probabilities, hence separate from those that underlie the Ellsberg paradox.

- Allowing preferences over final wealth distributions that are nonlinear in the probabilities doesn't help: Safra and Segal (2009 *Econometrica*) show that the reactions are separate from those that underlie the Allais paradox.

- And kinks can't be ubiquitous enough to fix expected-utility-of-wealth Maximization's mispredictions of aversion to small risks because a concave vN-M utility function must be differentiable almost everywhere.

End of aside

Background on reference-dependence

Kahneman and Tversky (1979 *Econometrica*) stress that prospect theory's emphasis on changes from reference points is a basic aspect of human nature:

> An essential feature of the present theory is that the carriers of value are changes in wealth or welfare, rather than final states. This assumption is compatible with basic principles of perception and judgment. Our perceptual apparatus is attuned to the evaluation of changes or differences rather than to the evaluation of absolute magnitudes. When we respond to attributes such as brightness, loudness, or temperature, the past and present context of experience defines an adaptation level, or reference point, and stimuli are perceived in relation to this reference point (Helson (1964)). Thus, an object at a given temperature may be experienced as hot or cold to the touch depending on the temperature to which one has adapted. The same principle applies to non-sensory attributes such as health, prestige, and wealth. The same level of wealth, for example, may imply abject poverty for one person and great riches for another depending on their current assets.

Kahneman (December 2003 *AER*) gives powerful visual examples, whose illusions persist after they are pointed out and their mechanisms understood):

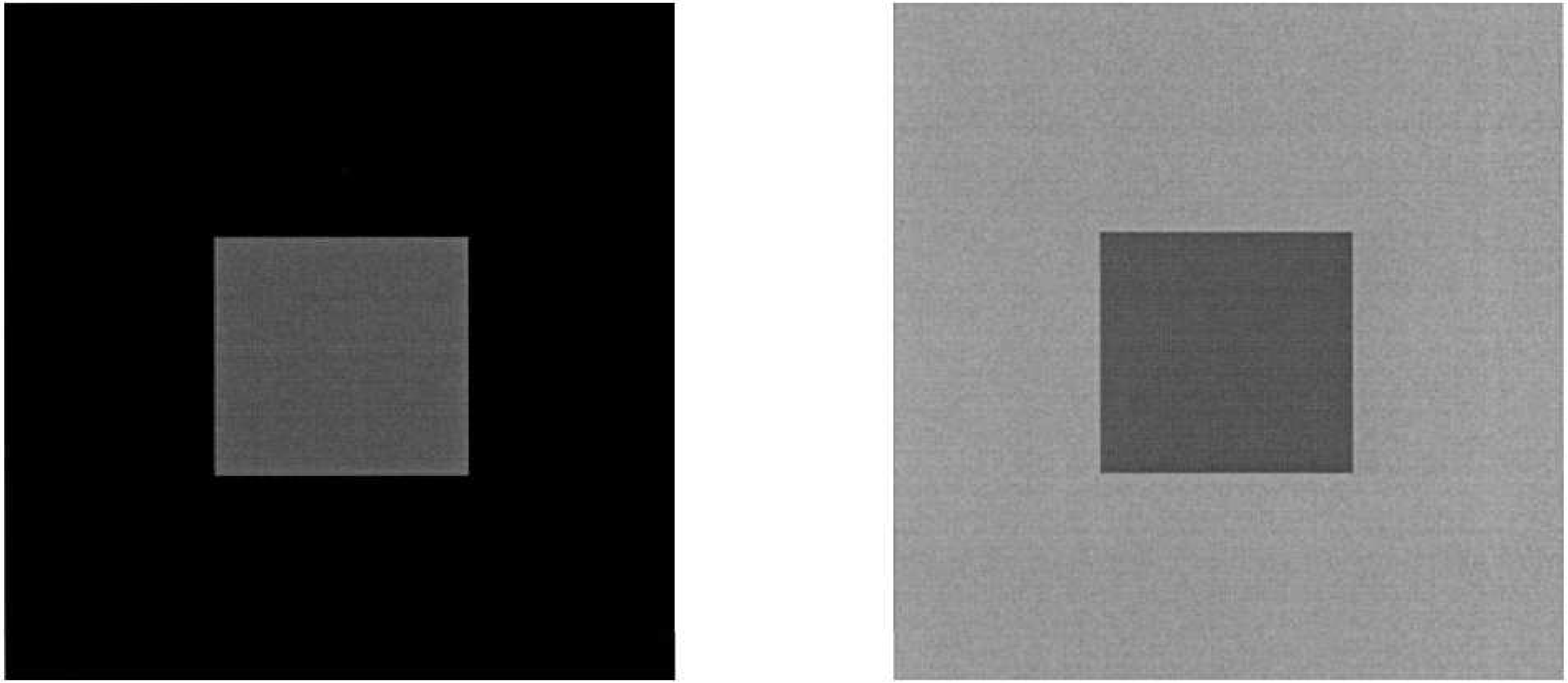

FIGURE 5. REFERENCE-DEPENDENCE IN THE PERCEPTION OF BRIGHTNESS

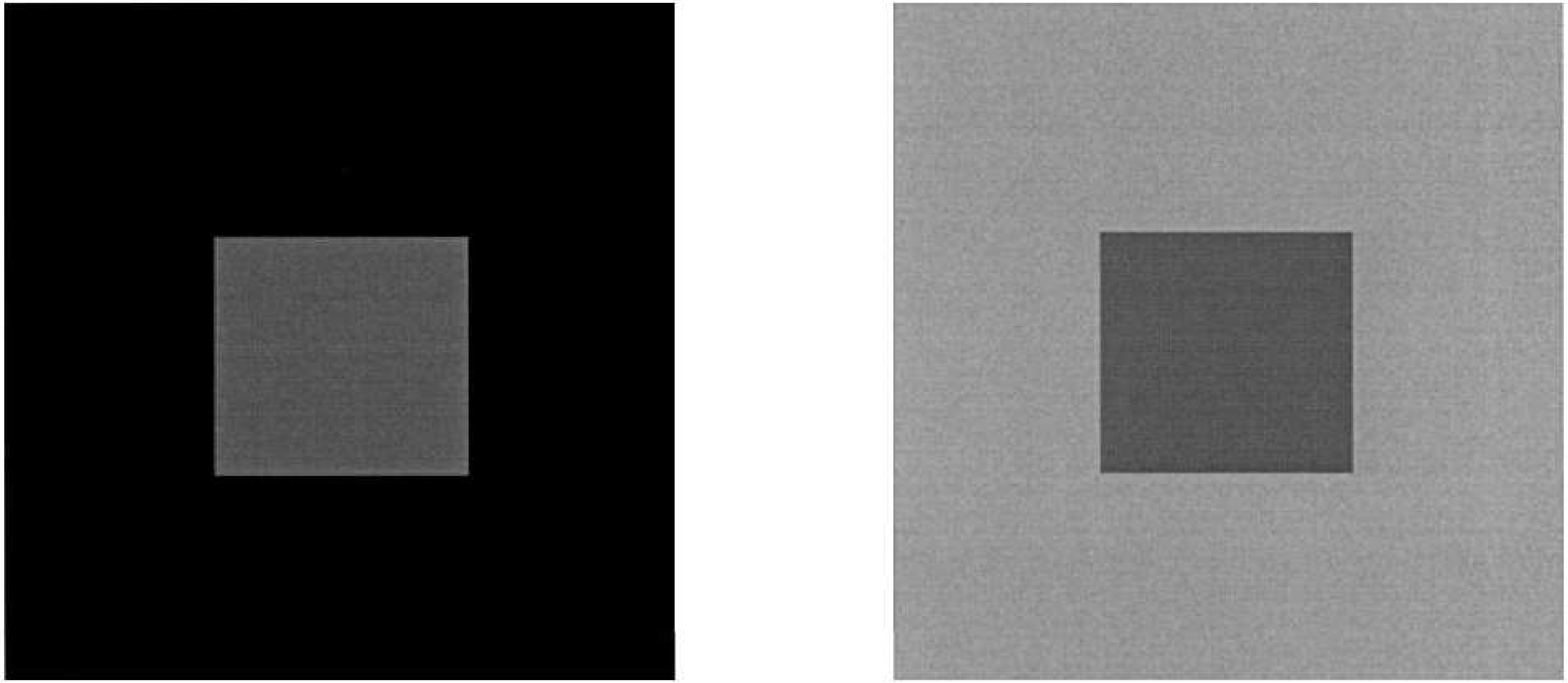

FIGURE 5. REFERENCE-DEPENDENCE IN THE PERCEPTION OF BRIGHTNESS

(The two inner squares are equally bright.)

FIGURE 7. AN ILLUSION OF ATTRIBUTE SUBSTITUTION

FIGURE 7. AN ILLUSION OF ATTRIBUTE SUBSTITUTION

(The two horsies are exactly the same size.)

The three main elements of prospect theory

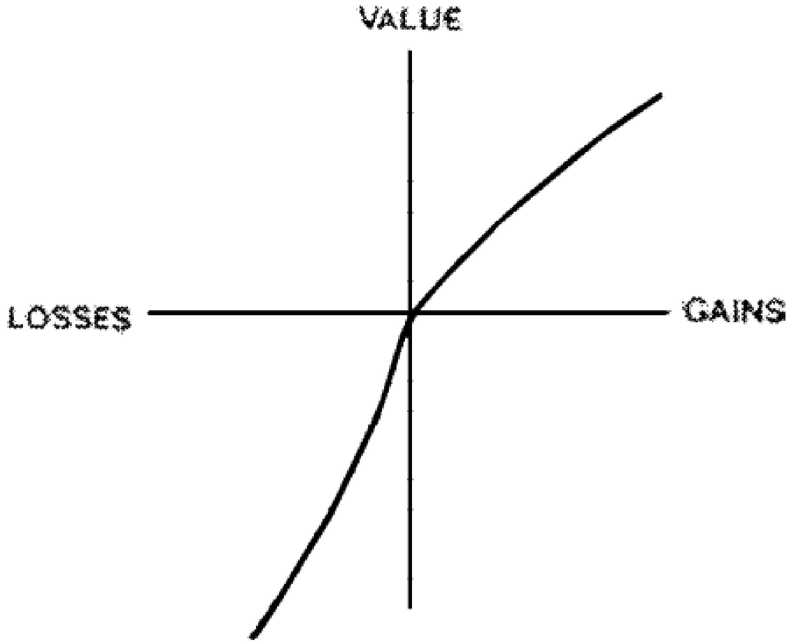

FIGURE 6. A SCHEMATIC VALUE FUNCTION FOR CHANGES

● Loss aversion

The kink at 0, the reference point, means a small decrease below it hurts more than an equal increase above it helps. The "coefficient of loss aversion" is the ratio of marginal value loss below to marginal value gain above, empirically around 2 or 3, never less than 1. It does vary a bit from person to person, and from context to context—is it more painful to lose an apple or a banana? On Tuesday or Friday? But it's remarkably stable for an empirical parameter.)

● **Diminishing sensitivity**

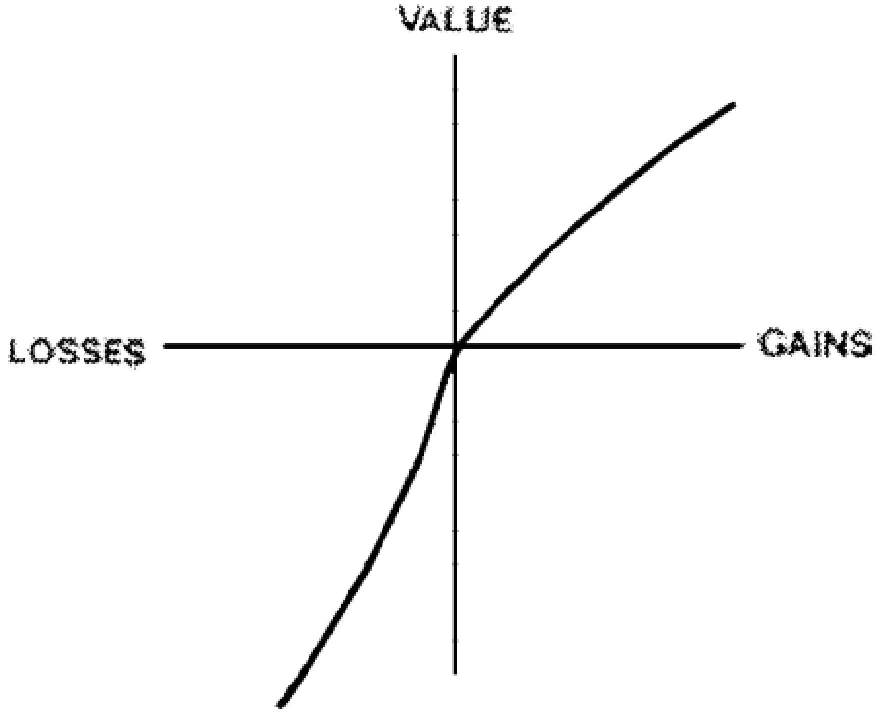

FIGURE 6. A SCHEMATIC VALUE FUNCTION FOR CHANGES

The value function exhibits diminishing marginal sensitivity to losses as well as gains, making it concave for gains but convex for losses.

Kahneman and Tversky (1979 *Econometrica*) argue that diminishing sensitivity reflects a fundamental feature of human cognition and motivation:

> Many sensory and perceptual dimensions share the property that the psychological response is a concave function of the magnitude of physical change. For example, it is easier to discriminate between a change of 3 and a change of 6 in room temperature, than it is to discriminate between a change of 13 and a change of 16. We propose that this principle applies in particular to the evaluation of monetary changes…. Thus, we hypothesize that the value function for changes of wealth is normally concave above the reference point ... and often convex below it....

But diminishing sensitivity is more relevant to decisions under uncertainty, so I don’t focus on it here.

● **Nonlinear probability weighting (can't be seen in the picture!)**

A third feature of prospect theory, nonlinear probability weighting, is a kind of fudge factor by which people are assumed to overweight small probabilities and underweight large ones, so that the value of a risk is $\pi(p)v(x) + \pi(q)v(y)$ rather than $pv(x) + qv(y)$.

Nonlinear probability weighting is less important and less well established empirically than loss aversion and diminishing sensitivity, and will not be discussed here even when we consider decisions under uncertainty.

Although the behavioural literature sometimes makes a big deal about diminishing sensitivity and even nonlinear probability weighting, and they are realistic and important for some applications, most of the action in prospect theory comes from reference-dependence and loss aversion.

I will focus on those aspects here.

Return of the mug people

Recall that KKT randomly gave mugs to half the subjects (“owners”) and nothing to the others (“nonowners”), then elicited selling prices for owners and buying prices for non-owners.

Supply and demand “should” be mirror images of each other. But the average buying price of non-owners was about \$3.50, while the average selling price of owners was about \$7.00: Way too big a gap to be random.

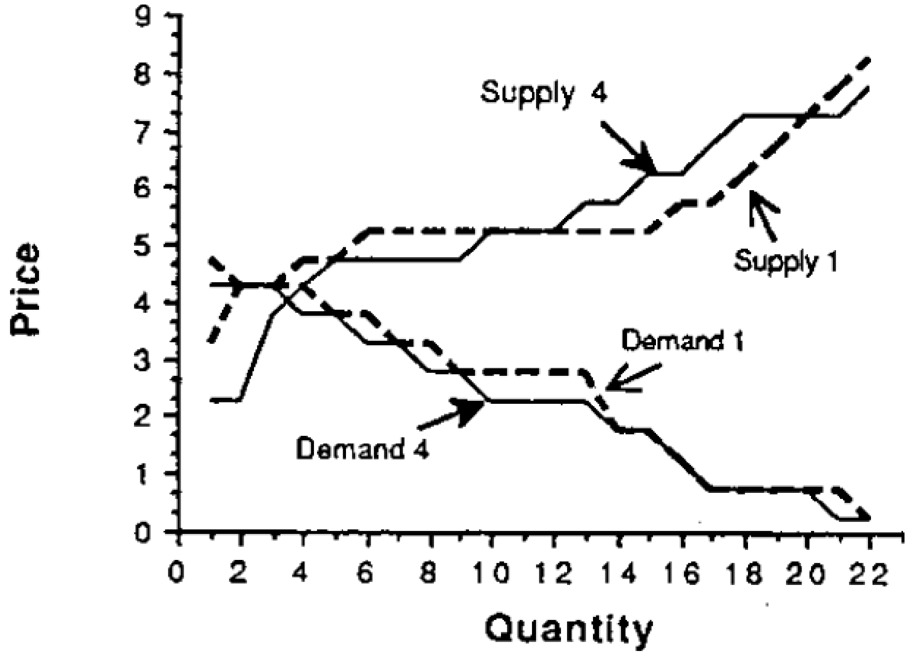

Fig. 1.—Supply and demand curves, markets 1 and 4

How do we model this with reference-dependence and loss aversion?

It seems natural to assume that subjects consider their mug-money choices in isolation, without trying to integrate them into a lifetime consumption plan.

But in other applications it is necessary to think harder about how people group choices in thinking about them: “mental accounting” and “bracketing”.

Imagine (like Kőszegi and Rabin 2006 *QJE*) that people have both ordinary consumption and “gain-loss” utilities for both mugs and money.

Assume that subjects have linear consumption utility: value = value of mug (or not) + money, and owners’ and nonowners’ consumption utilities for mugs are uniformly distributed between $0 and $9 (using the natural metric).

Assume subjects also have gain-loss utilities, with no diminishing sensitivity but with a coefficient of loss aversion of 2, so that losses relative to the reference point lower their gain-loss utility twice as much as gains raise it.

The relative weight of gain-loss utility is η, so total utility is consumption utility + $\eta \times$ gain-loss utility.

Subjects' reference points are determined by their expectations.

Suppose first that:

- Owners expect to keep their mugs (and gain no money).

- Nonowners expect to keep their money (and gain no mug).

Supply of mugs

An owner with mug consumption value $\$v$ who is considering trading his mug for $\$m$ will compare his total (consumption plus gain-loss) utility from keeping his mug with her/his total utility from trading the mug for $\$m$.

Because as an owner he expected to keep her/his mug, if he keeps it there are no gain-loss surprises on the mug or money dimension. His total utility of keeping = consumption utility (v + 0) + η × gain-loss utility (0 + 0).

If he trades his mug for $\$m$, there are gain-loss surprises on both dimensions, “losing” him $\eta \times 2v$ on the mug dimension—because it’s *his* mug, and the coefficient of loss aversion is 2—but gaining him $\eta \times m$ on the money dimension—only m, because it’s *someone else*’s money. His total utility from trading = consumption utility (0 + m) + η × gain-loss utility ($-2v + m$).

Thus the lowest price m at which he would be willing to sell his mug is the lowest m that makes $v \leq m + \eta(-2v + m)$, or

$$m^* = v(1 + 2\eta)/(1 + \eta).$$

If $\eta = 0$ we get the usual $m^* = v$ result. But if $\eta > 0$, say $\eta = 1$, we get $m^* = 1.5v$, which yields average selling price \$6.75 ≈ KKT’s \$7.

(A distribution makes it easy to generate an entire supply curve as above.)

Demand for mugs

Similarly, a non-owner with mug consumption value \$*v* who is considering trading \$*m* of his money for a mug will compare his total utility from keeping her/his \$*m* with his total utility from trading \$*m* for a mug.

Because as a nonowner he expected to keep his \$*m*, if he keeps it there are no gain-loss surprises on either the money or the mug dimension. His total utility of keeping = consumption utility $(0 + m)$ + $\eta \times$ gain-loss utility $(0 + 0)$.

If he trades her/his \$*m* for a mug, there are gain-loss surprises on both dimensions, gaining him $\eta \times v$ on the mug dimension but losing him $\eta \times 2m$ on the money dimension. His total utility from trading = consumption utility $(v + 0)$ + $\eta \times$ gain-loss utility $(v-2m)$.

Thus the highest price $\hat{m}$ he would be willing to pay for the mug is the highest *m* that makes

$v + \eta(v - 2m) \geq m$, or $\hat{m} = v(1 + \eta)/(1 + 2\eta)$.

If $\eta = 0$ we get the usual $\hat{m} = v$ result; but if $\eta > 0$, say $\eta = 1$, we get $\hat{m} = 0.67v$, which yields an average buying price of \$3.00 ≈ KKT's \$3.50.

(A distribution makes it easy to generate an entire demand curve as above.)

This explanation links two widespread empirical regularities, the prevalence of gaps with WTA > WTP and the prevalence of loss over gain aversion.

To better understand the role of expectations, re-do the above analysis, with $\eta = 1$, for a mug-owner who expects to sell his mug, say for \$*x* (so his reference point is having \$*x* and no mug). Then re-do it for a nonowner who expects to buy a mug for \$*y* (so his reference point is having a mug but -\$*y*).

These expectations make sellers and buyers more willing to trade, creating a preference bias relative to the standard model, in favor of the expected.

That is the reasoning behind this Kőszegi and Rabin quotation:

> Our theory…supports the common view that the "endowment effect" found in the laboratory, whereby random owners value an object more than nonowners, is due to loss aversion…. But our theory makes the less common prediction that the endowment effect among such owners and nonowners with no predisposition to trade will disappear among sellers and buyers in real-world markets who expect to trade….

Re-enter cabdrivers

Recall that to explain their results for cabdrivers' labor supply, Camerer et al. informally proposed a model in which drivers have daily income targets and work until the target is reached.

They therefore tend to work less on days when "wages" are high.

In the proposed explanation, the reference point is a daily income target.

Loss aversion creates a kink that tends to make realized income respond to the income target as well as the wage, and bunch around the target.

But Farber argued, treating the income target, which is not directly observed, as a latent variable, that the target is too unstable and imprecisely estimated to yield a useful reference-dependent model of labor supply.

Motivated in part by Camerer et al.’s and Farber’s analyses, Kőszegi and Rabin (2006 *QJE*) proposed an implementation of prospect theory that is more general than Farber’s income-targeting model in most respects but takes a more specific position on how targets are determined.

In their theory as applied to cabdrivers’ labor supply:

- A driver’s preferences reflect both the standard consumption utility of income and leisure and reference-dependent “gain-loss” utility, with their relative importance tuned by an estimated parameter.

- A driver has a daily target for hours as well as income, and as in Farber’s model he is loss-averse, but working longer than the hours target is now a loss, just as earning less than the income target is.

- Most importantly, the targets are endogenized by setting them equal to a driver’s theoretical rational expectations of hours and income (Kőszegi and Rabin’s notion of “preferred personal equilibrium”).

Abeler et al. (2011 *AER*) conducted a careful experimental test of Kőszegi and Rabin’s expectational view of reference points, and largely confirmed it.

Kőszegi and Rabin's treatment of the targets as rational expectations and their model's distinction between the effects of anticipated and unanticipated wage increases has the potential to reconcile:

- The negative wage elasticity of hours found by Camerer et al. (1997 *QJE*) and Farber (2005 *JPE*, 2008 *AER*); and

- The positive relationships between *expected* earnings and labor supply found by Oettinger (1999 *JPE*), Fehr and Goette (2007 *AER*), and others. In their model reference-dependence has no effect when expectations are exactly realized, in which case the model reduces to its neoclassical part.

Crawford and Meng (2011 *AER*) then used Farber’s data to estimate a model based on Kőszegi and Rabin’s theory.

They followed Kőszegi and Rabin in dropping diminishing sensitivity and nonlinear probability weighting (for both of which there is evidence, but less than for reference-dependence and loss aversion).

Thus their model is fully consistent with rationality, with concave objective functions. Its only important deviation from a neoclassical model is adding targets to income and leisure in the domain of preferences.

They also assumed for simplicity that the targets are point expectations rather than distributions as in Kőszegi and Rabin’s theory.

Further, in the structural estimation that parallels Farber’s (2008) analysis, they allowed for consumption as well as gain-loss utility and hours as well as income targets as Kőszegi and Rabin’s theory suggests.

Crawford and Meng followed Farber's (2005 *JPE*, 2008 *AER*) econometric strategies, but treated targets as rational expectations, not latent variables.

They operationalized the targets by finding natural sample proxies for rational expectations with limited endogeneity problems.

Recall that in Farber's dataset estimating the targets as latent variables caused computational problems, which were what led him to conclude that the income targets in his model are too unstable and imprecisely estimated to yield a useful reference-dependent model of labor supply.

The additional structure from treating the targets as proxied rational expectations avoids some of the problems Farber encountered, and allows tests of the model by looking for systematic shifts in drivers' stopping decisions associated with the targets.

Crawford and Meng showed that their estimated model can:

- Reconcile the negative wage elasticity of hours found by Camerer et al. and Farber with the positive relationships between *expected* earnings and labor supply found by Oettinger, Fehr and Goette, and others.

- Reconcile the smoothness of the relationship between stopping probability and realized income Farber found.

And (despite Farber's negative conclusion) it can:

- Yield estimates of the targets that are stable and sufficiently precisely estimated to yield a useful reference-dependent model of labor supply.

Model details

Treating each day separately as in all previous analyses, consider the preferences of a given driver during his shift on a given day.

I and *H* denote his income earned and hours worked that day.

I^r and H^r denote his income and hours targets for the day.

His total utility, $V(I, H|I^r,H^r)$, is a weighted average of consumption utility $U_1(I) + U_2(H)$ and gain-loss utility $R(I, H|I^r,H^r)$, with weights $1 – \eta$ and η ($0 \leq \eta \leq 1$):

(1) $$V(I,H \mid I^r,H^r) = (1-\eta)(U_1(I)+U_2(H)) + \eta R(I,H \mid I^r,H^r),$$

where gain-loss utility

(2) $$R(I,H \mid I^r,H^r) = 1_{(I-I^r \leq 0)}\lambda(U_1(I)-U_1(I^r)) + 1_{(I-I^r>0)}(U_1(I)-U_1(I^r)) + 1_{(H-H^r \geq 0)}\lambda(U_2(H)-U_2(H^r)) + 1_{(H-H^r<0)}(U_2(H)-U_2(H^r)).$$

(1)-(2) incorporate several of Kőszegi and Rabin's provisional assumptions:

- Consumption utility is additively separable across income and hours, with $U_1(\cdot)$ increasing in I, $U_2(\cdot)$ decreasing in H, and both concave.

- Gain-loss utility is also separable, determined component by component by differences between realized and target consumption utilities.

- Gain-loss utility is a linear function of those utility differences, ruling out Prospect Theory's "diminishing sensitivity" as in a leading case Kőszegi and Rabin sometimes focus on (their Assumption A3').

- Losses have a constant weight λ relative to gains, "the coefficient of loss aversion," which is the same for income and hours. Empirically, $\lambda \approx 2$ to 3.

(1)-(2) depart from Kőszegi and Rabin in treating drivers' targets as deterministic point expectations, a natural simplification given that the model (unlike theirs) makes explicit allowance for errors and therefore can have gains and losses even with point expectations.

(This may exaggerate the effect of loss aversion, and if anything it biases the comparison against a reference-dependent model and in favor of a neoclassical model.)

Crawford and Meng follow Kőszegi and Rabin in equating the income and hours targets I^r and H^r to drivers' rational expectations, with natural sample proxies.

If gain-loss utility has small weight, Kőszegi and Rabin's model approaches a neoclassical model, with standard implications for labor supply.

Even when gain-loss utility has large weight, the standard implications carry over for changes in the wage that are perfectly anticipated.

But when realized wages deviate from expected, the probability of stopping may be more strongly influenced by hours or income, depending on which target is reached first, and the model deviates from a neoclassical model.

On a good day (wage higher than expected) a driver hits the income target before the hours target; on a bad day a driver hits the hours target first.

Crawford and Meng's structural estimates suggest that most drivers stop near the second target they reach on a given day: hours on a good day, income on a bad day.

Given that $\lambda \geq 1$ the model allows a simple characterization of a driver's optimal stopping decision with targets for hours as well as income.

Suppose that a driver expected the wage to remain constant at w^e.

Then his optimal stopping decision maximizes reference-dependent utility $V(I, H|I^r,H^r)$ as in (1) and (2), subject to the linear menu of income-hours combinations $I = w^e H$.

When $U_1(\cdot)$ and $U_2(\cdot)$ are concave, $V(I, H|I^r, H^r)$ is concave in I and H for any given targets I^r and H^r. (This depends on ruling out "diminishing sensitivity".)

Thus the driver's decision is characterized by a first-order condition, generalized to allow kinks at the reference points: He continues if and only if the anticipated wage w^e exceeds the relevant marginal rate of substitution.

Table 1 lists the marginal rates of substitution in the four possible gain-loss regions, expressed as hours disutility costs of an additional unit of income.

(On boundaries, marginal rates of substitution are replaced by generalized derivatives whose left- and right-hand values equal the interior values.)

Table 1. Marginal Rates of Substitution with Reference-Dependent Preferences

	Hours gain ($H < H^r$)	**Hours loss ($H > H^r$)**
Income gain ($I > I^r$)	$-U_2'(H)/U_1'(I)$	$-[U_2'(H)/U_1'(I)][1-\eta+\eta\lambda]$
Income loss ($I < I^r$)	$-[U_2'(H)/U_1'(I)]/[1-\eta+\eta\lambda]$	$-U_2'(H)/U_1'(I)$

When hours and income are both in the gains or loss domain, the marginal rate of substitution is the same as for consumption utilities alone, so the stopping decision satisfies the standard neoclassical first-order condition.

When hours and income are in opposite domains, the marginal rate of substitution equals the consumption-utility trade-off times either $(1 - \eta + \eta\lambda)$ (> 1 when $\lambda > 1$) or $1/(1 - \eta + \eta\lambda)$.
(The tradeoff favors work more than the neoclassical tradeoff in the income loss/hours gain domain, but less in the hours loss/income gain domain.)

Figure 1 illustrates the driver's optimal stopping decision when the wage is higher than expected, so the income target is reached before hours target.

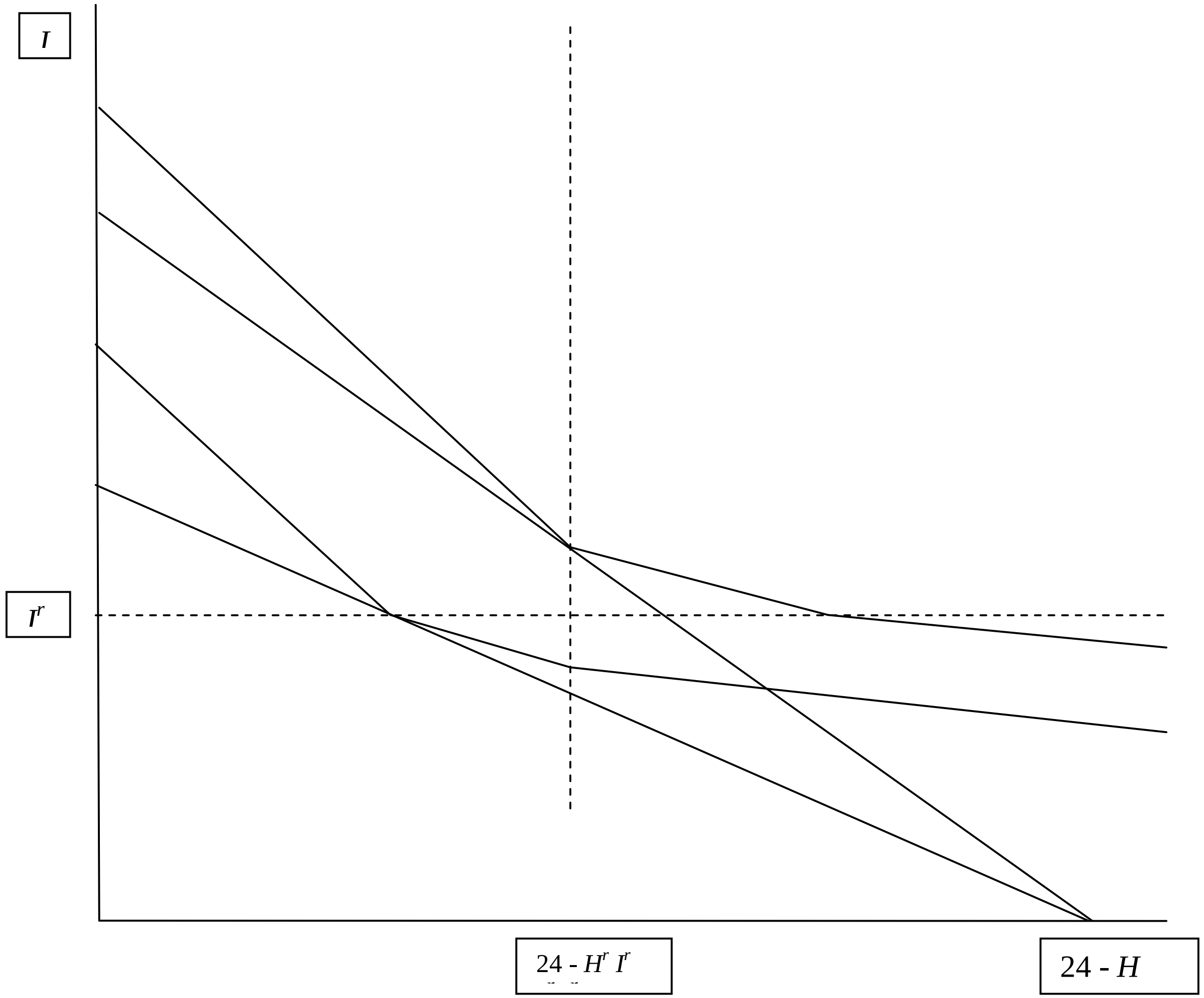

Figure 1: A Reference-dependent Driver's Stopping Decision

Letting I_t and H_t denote income earned and hours worked by the end of trip t, the driver starts in the lower right-hand corner, with $(H_0, I_0) = (0, 0)$, and anticipates moving along a sample line $I = w^e H$.

A driver anticipates heading northwest along a random but monotone path, which is approximately continuous (the average trip length is 12 minutes).

He anticipates passing through a series of domains such that the hours disutility cost of income weakly increases as hours and income accumulate, reflecting the concavity of reference-dependent utility in I and H.

In the initial, income-loss/hours-gain domain (Table 1, Figure 1), the comparison between the anticipated wage and the hours disutility cost favors working more than the neoclassical comparison would.

But for a given anticipated wage the tradeoff becomes (weakly) less and less favorable as income and hours accumulate.

He continues working as long as the anticipated wage w^e exceeds the hours disutility cost of an additional unit of income. Given the assumptions about his expectations, his decision appears globally optimal to him.

Figure 2 compares the labor-supply curves for a neoclassical driver and a reference-dependent driver with the same consumption utility functions.

The solid curve is the neoclassical supply curve, while the dashed curve is the reference-dependent one.

The shape of the curve depends on which target has a larger influence on the stopping decision, which depends in turn on the relationship between the neoclassical optimal stopping point (that is, the stopping point that maximizes consumption utility alone) and the targets.

Figure 2 illustrates the case suggested by Crawford and Meng's estimates:

For wages that reconcile the income and hours targets as at point D, the neoclassically optimal income and hours are higher than the targets, so the driver stops at his second-reached target.

Whenever the wage is to the left of point D, the hours target is reached before the income target, and vice versa.

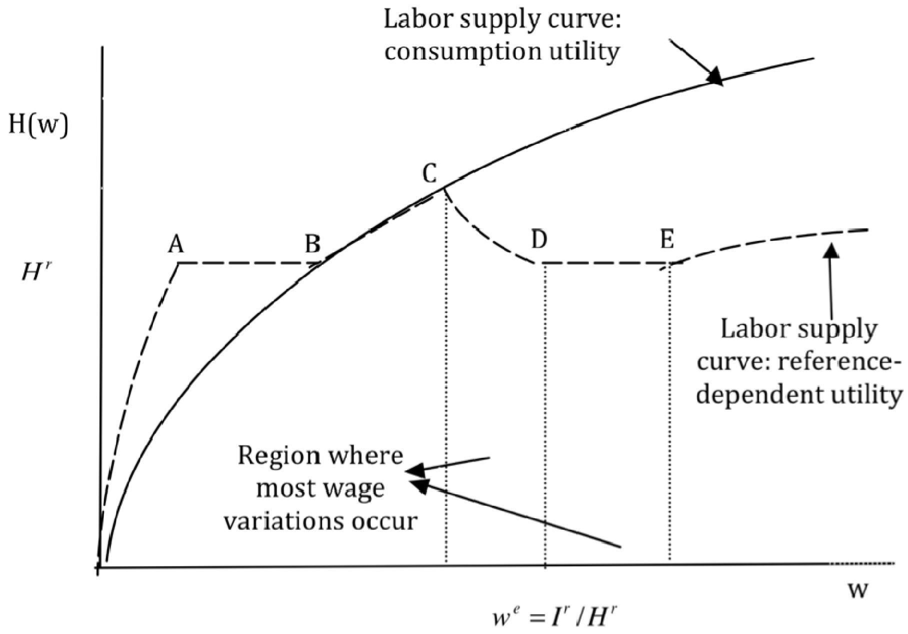

Figure 2: A Reference-dependent Driver’s Labor Supply Curve

As Figure 2 illustrates, reference-dependent labor supply is non-monotonic.

When the wage is to the left of point A, the higher cost of income losses raises the incentive to work above its neoclassical level.

Along segment AB labor supply is determined by the kink at the hours target, which is reached first.

Along segment BC the neoclassical optimal stopping point is above the hours but below the income target, so the gain-loss effects cancel out, and reference-dependent and neoclassical labor supply coincide.

Along segment CD labor supply is determined by the kink at the income target, which is reached second, so the wage elasticity of hours is negative.

Along segment DE labor supply is determined by the kink at the hours target, which is reached second.

Finally, when the wage is to the right of point E, the higher cost of hours losses lowers the incentive to work below its neoclassical level.

Most realized wages fall close to point D.

Recall that Crawford and Meng's structural estimates suggest that most drivers stop near the second target they reach on a given day.

Whenever the income target has an important influence on a driver's stopping decision, even a driver who values income but is "rational" in the sense of prospect theory has a negative wage elasticity of hours.

But when the hours target is dominant, the elasticity is near zero.

The aggregate elasticity is negative as Camerer et al. and Farber found.

These effects are driven by transitory wage changes, but anticipated wage changes have the same effect on labor supply as in a neoclassical model.

Further, the heterogeneity of realized wages yields a smooth aggregate relationship between stopping probability and realized income.

Thus, the model reconciles Farber's finding that stopping probabilities are significantly related to hours but not income with a negative aggregate wage elasticity of hours and a positive elasticity for anticipated wage changes.