These are slides from Matthew's lectures for the AEA continuing education program in behavioral economics, Atlanta, January 5-7. They are meant to be supplements for those attending the lectures (and listening to the clarifications, caveats, and discussion), not as stand-alone documents.

# "Behavioral Economics"

Matthew Rabin (Today & Tomorrow)

University of California — Berkeley

David Laibson (Tomorrow & Thursday)

University of Harvard — Cambridge

These lectures: research introducing psychologically more realistic assumptions into economics.

Brief introduction to some themes, findings, insights, and implications of behavioral economics, with emphasis on topics conducive to formal modeling and empirical testing.

# A "Wave Theory" of Behavioral Economics

First Wave: Identify "anomalies" — ways that economic theory has been importantly wrong, and identify some alternative conceptualizations.

Second Wave: Formalize some of the alternatives in precise models, and identify some empirical validations of these models.

Third Wave: *Fully* integrate into economic analysis by embedding old and new assumptions as special cases of general models, formulate new theoretical results, empirical tests, and *applications* in entirely normal-science ways, with attention to all the desiderata of economic models (parsimony, predictiveness, generality, relevance, insight, etc.)

We've now entered the $3^{rd}$ Wave.

Two premises of third-wave behavioral economics

Premise 1: Adding untraditional assumptions doesn't at all mean abandoning traditional methods.

Premise 2: Adding untraditional assumptions doesn't mean abandoning traditional assumptions.

Goal not just "relaxing" restrictions. Aspire to modifications, improvements.

Tiny section of what could be included. And heavy emphasis on the 'positive' and shovel-ready. Also heavy emphasis on themes, illustrative evidence, and conceptualization of the improvements.

My lectures reflect: a) severely limited time to present material, and b) my prejudices and tastes.

Categories of Improvement

More realistic utility functions.

"Errors": Superhuman rationality → human rationality.

## Topic 2: New Models of Preferences (Matthew)

Reference-dependence, social preferences

## Topic 3: Quasi-Utility Maximization (Matthew)

Narrow bracketing, misprediction of preferences

## Topic 4: Present-Biased Preferences, a.k.a. Hyperbolic Discounting (Matthew & David)

## Topic 5: Applications (David)

## Topic 6: Normative and Policy Implications (David)

Plenty of prominent and important topics missing.

E.g.

## Heuristics and Biases

Mr. Burns: I need someone to take care of my house. Who's that guy who screws up everything?

Smithers: Simpson, sir.

Mr. Burns: I'll ask him. The way I figure, he's due for a good performance.

## Behavioral Finance

Homer: "This year I invested in pumpkins. They've been going up the whole month of October and I got a feeling they're going to peak right around January. Then bang! That's when I'll cash in!"

# Overview Readings

"First Wave": Kahneman and Tversky, eds., *Choices, Values, and Frames*, 2000, Richard Thaler: "Anomalies" columns in Journal of Economic Perspectives (Collected in the book *The Winner's Curse*)

"2nd Wave": Rabin, Matthew, "Psychology and Economics" *JEL* 1998; Camerer, Loewenstein, and Rabin, eds. 2003, "Advances in Behavioral Economics"

"3rd Wave": Stefano DellaVigna: "Psychology and Economics: Evidence from the Field", *JEL* 2009.

# Some Psychological Principles

(heretofore underappreciated by economists)

1. Reference dependence and contrasts

2. Proportional thinking

3. Non-integration, non-bracketing

4. Failure of contingent thinking

5. Salience/attention/focusing effects

6. Ego/Identify/Image as motivation

7. Limits to human rationality not *just* about 'complexity' of, or benefits to, getting things right.

# Some Psychological Principles

(heretofore underappreciated by *psychologists*)

1. When the cost/price of some activity goes up, people do less of it.

2. People make tradeoffs.

3. Observed behavior is not a *direct* manifestation of intrinsic preferences.

# Implications of Principles

(heretofore underappreciated by economists)

Underappreciated by economists: Details, not just gestalt costs and benefits, can matter a lot.

Corollary: "Life courses" are not life plans.

**Homer: "I bought this [book of classics] when Bart was born—I was going to read to him every day."**

**Bart: "You've never read to me! What happened?"**

**Homer: "Stuff came up. Mostly car related."**

Underappreciated by psychologists: equilibrium

# Ways greater psychological realism can improve economic analysis:

1. Explaining behavior studied by economists that traditional analysis has had difficulties explaining.

2. Explaining behavior that seems economically important enough that one would have *thought* economists would have been studying—but haven't been.

3. Beyond explaining *behavior*, better understand normative/hedonic effects of observed behavior.

4. Often by making our models more complicated and less tractable but more realistic in a trade off *on the scale of what economists do all the time* (when invoking more familiar assumptions)

Atlanta2.tex

January, 2010

# More Realistic Preferences

Homer: "All my life I've had one dream: to achieve my many goals."

# “De Gustibus” for 21st Century

istorically and currently, many of the newer assumptions from BE are an attempt to incorporate more realistic assumptions about preferences ... often *rather* than extreme-irrationality assumptions that appeal to other researchers.

Shouldn’t cling to bad rational explanations for choices just to keep irrationality assumptions out of economic analysis.

But *nor* should we cling to “classical” assumptions about preferences, and hence automatically assume “anomalous” behaviors are just mistakes.

De gustibus non est disputandum
... exceptum if non-selfishum
De gustibus non est disputandum
... exceptum if non-consequentialum
De gustibus non est disputandum
... exceptum if belief-basedum
De gustibus non est disputandum
... exceptum if non-recursivum

In cognitively simple binary choices with clear consequences, choice reflects preference.

Most people reject 50/50 gain $40, lose $35 bets. And many buy extended warranties, etc.

Now review preferences that explain this. Tomorrow: wonder if mistakes also involved.

But it is *not* a mistake that people with reference-independent utility would make.

# Reference Dependence

Moe: "If you want to signal me, use this bird call."

*[Moe whistles like a bird. An eagle swoops down and pecks him on the face]*

"Ow! Not the face!"

*[the eagle switches to pecking Moe in the groin]*

"Ooh! Ooh! Okay, the face!"

*[the eagle switches back]*

"Ooh! Whoa, that actually feels good after the crotch!"

Human perceptions, feelings, decisions, etc. are all determined very largely by comparisons to reference levels rather than by absolute levels.

In virtually all physiological and psychological reactions to (e.g. temperatures), people's reactions tend to reflect adaptation, change, and contrast, rather than solely absolute levels of outcomes.

Two main features of reference dependence emphasized by Kahneman and Tversky and (we) acolytes:

## 1. Loss Aversion

Kahneman, Knetsch, and Thaler (+ dozens of replications and a couple very determined non-replications):

If randomly distribute mugs, designate some people owners and others non-owners, the elicit buying/selling/choosing prices, you get:

Buying (+ choosing) price $\approx$ \$3.50
Selling price $\approx$ \$7.00

More important (and more robust):

Accept 50/50 lose \$600, gain \$700 bet? Vast majority of people turn this down.

But it is not because of DMU(W).

Aversion to modest scale risk cannot come from DMU(W). Strongest such aversion involve mix gains and losses relative to status quo—and derive from loss aversion.

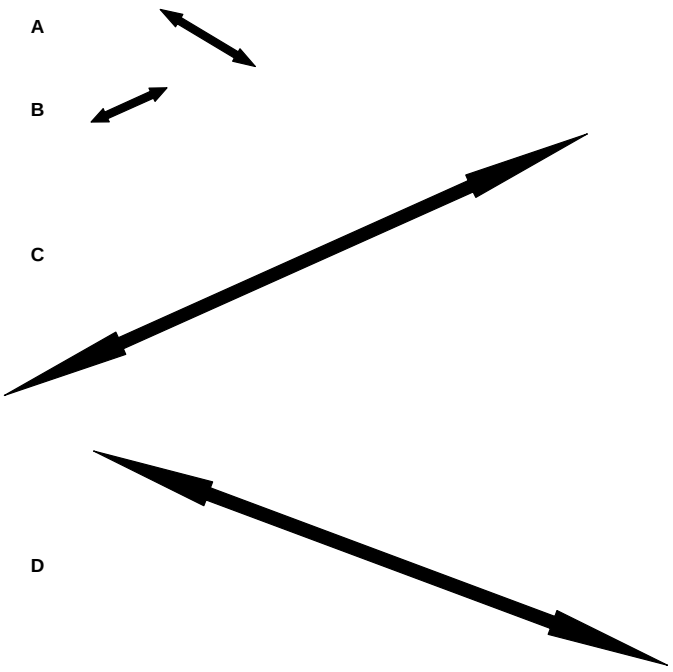
Other domains where LA important:

Moral/Fairness
Hippocratic Oath
Wage Cuts

## 2. Diminishing Sensitivity

Another meta fact about human psychology: Across domains, we tend to perceive, judge, and choose based on proportional thinking.

Which of the following 4 lines is shortest? Longest?



For each of the following pairs, which “feels” like a bigger difference?

visually 88 ft. away vs. 87 ft. away

vs.

2 ft. vs. 1 ft.

gratification 87 days from now vs. 88 days

vs.

gratification 0 days vs. 1 day

saving $10 on $1,000 item

vs.

saving $10 on $20 item

carrying a suitcase 21 blocks vs. 20 blocks

vs.

4 blocks vs. 3 block

19% chance vs. 18% chance

vs.

1% chance vs. 0%

vs.

100% chance vs. 99%

18 heads/13 tails vs. 16 heads/15 tails

vs.

5 heads/0 tails vs. 3 heads/2 tail

gaining $88 vs. gaining $87

vs.

gaining $2 vs. gaining $1

vs.

gaining $1 vs. gaining $0

losing $88 vs. losing $87

vs.

losing $2 vs. losing $1

vs.

losing $1 vs. losing $0

And if it holds for gaining and losing money ...

Prefer $420 for sure or 50/50 chance at $900?
Prefer losing $420 or 50/50 chance lose $900?

Often, of course, diminishing sensitivity is "normative" for consumption reasons.

Prefer \$4.2 million or 50/50 chance at 9 million?

Or DMRS in consumption

4 movies & 0 meals
2 movies & 2 meals
0 movies & 4 meals

But right thinking for consumption in other cases?

21st block is easier or harder than 4th block?

Value of dollar at $w - 900$, $w - 420$, $w$, $w + 420$, or $w + 900$?

Saving \$10 more valuable or less valuable when spending more?

Distance perception nothing to do with prefs ... probabilities? Proportions of coins? Invite departures from EU and Bayesian.

The psychology of gaining and losing does display diminishing sensitivity, but not for "instrumental" reasons.

# Modeling Reference-Dependent Utility

Utility not given by $u(c)$, but rather $u(c|r)$, where $c = (c_1, c_2, ...c_K)$ is consumption and $r = (r_1, r_2, ...r_K)$ a "reference level" of consumption. Fixing $r$, maximize expected value of $u(c|r)$.

**What is reference point?** Ignoring that ...

Suppose only concerned with preferences given a fixed reference level. Then can analyze choice in terms of people maximizing "value functions":

$$v(c|r) \equiv u(c|r) - u(r|r).$$

Dimension-by-dimension: for each $k$, define

$$v_k(c_k - r_k|r) \equiv$$
$$u(r_1, ..., c_k, ...r_K|r_1, ..., r_k, ...r_K)$$
$$-u(r_1, ..., r_k, ...r_K|r_1, ..., r_k, ...r_K).$$

Simplifying assumption of additive separability:

$$v(c|r) \equiv u(c|r) - u(r|r) = \sum_{k=1}^{K} v_k(c_k - r_k|r).$$

Letting $v(\cdot) \equiv v_k(\cdot|\cdot)$, consider Assumptions A0-A4 (inspired by Kahneman and Tversky, 79):

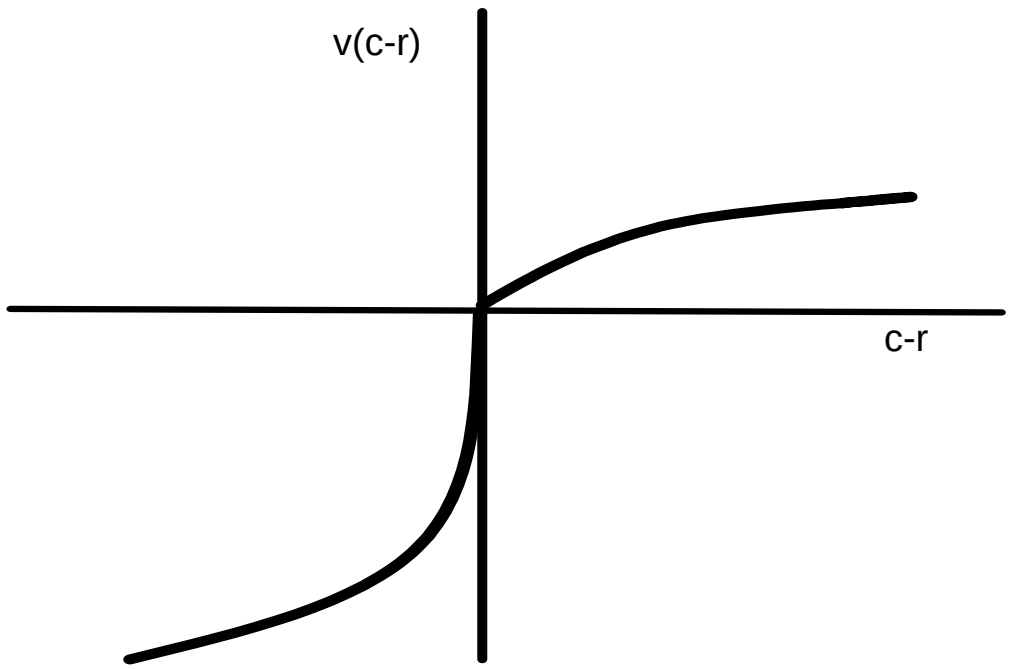
A0. $v(x)$ is continuous everywhere, twice differentiable for all $x \neq 0$, and $v(0) = 0$

A1. $v(x)$ is strictly increasing

A2. If $y > x > 0$, $v(y) + v(-y) < v(x) + v(-x)$.

A3. $v''(x) < 0$ for $x > 0$, $v''(x) > 0$ for $x < 0$.

A4. From $x > 0$ direction, $\lim_{x \to 0} \frac{v'(-x)}{v'(x)} \equiv L > 1$.



# Implications for Risk Attitudes?

1. Turn down any 50/50 lose $X/gain $X bets.

2. Risk averse among bets involving only gains

3. Risk-*loving* among bets involving only losses

4. "First-order risk-averse."

Implications 1 & 2 true of EU/DMU(W) model.

Implication 3 famously "new" implication of prospect theory inconsistent with EU/DMU(W), and verifiably true in important ways.

When ignoring DS: A3′. $v''(x) = 0$ for $x \neq 0$.

Implication 4: Folk wisdom that first-order risk aversion is *true*, but *inconsistent with* EU/DMU(W), and more important for economics

# DMU(W) & Modest-Scale Risk Aversion

The *standard* expected-utility model—utility defined over wealth (consumption) independent of reference levels—implies "second-order risk aversion":

But folk wisdom, formalized by Rabin (2000): not just limit result, but practical one. Without counterfactual (and insane) degree of concavity in $u(w)$, EU/DMU(W) predicts counterfactual (but sane!) virtual risk neutrality for modest stakes (e.g., \$1,000).

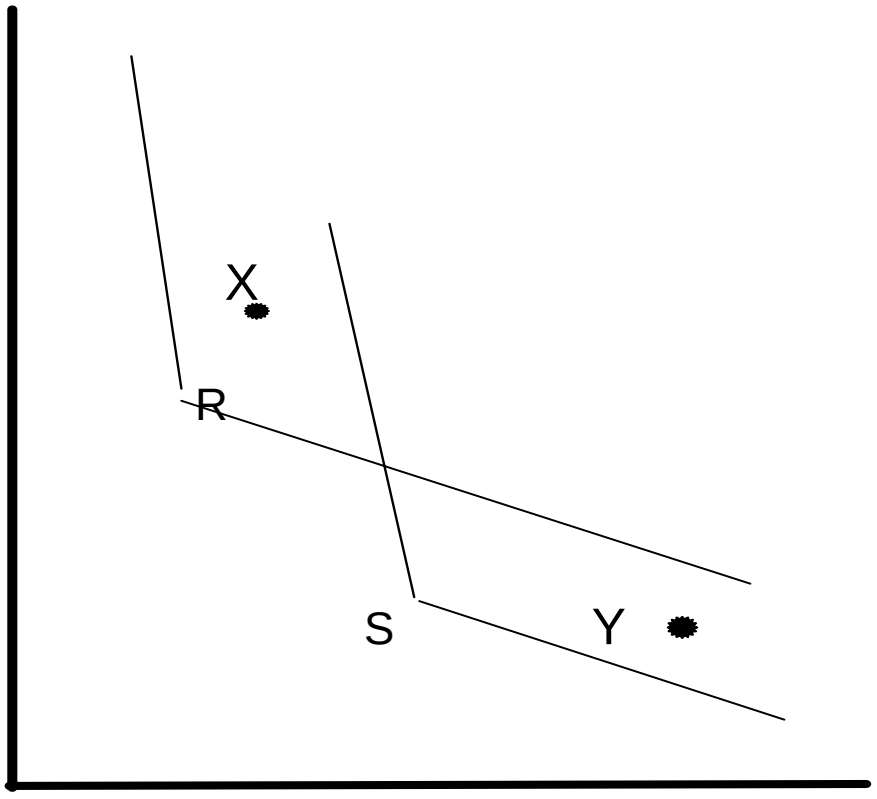
Explaining dislike of 50/50 gain \$11, lose \$10 bet by $u''(w) < 0$ says that \$21 more in lifetime wealth makes $u'(w)$ drop by 10%; if you were just \$168 wealthier, $u'(w) < (\frac{10}{11})^8 < \frac{1}{2}$ of current $u'(w)$! Not a chance. PDV(\$168) less than 10 cents a day ...

Problem is: people are *not* risk neutral over modest stakes, and this matters.

## What does this say in riskless choice?

Suppose that fixing a reference point, RD-VNM preferences over mugs and pens is given by $u(c_1, c_2|r_1, r$ $r_1 + r_2 + v(c_1 - r_1) + v(c_2 - r_2)$, where $v(x) = x$ for $x \geq 0$ and $v(x) = 2x$ for $x \leq 0$.

The reference points bundles R and S generate different indifference curves:




This generates endowment and status quo effects. E.g., $X \succ_R Y$ but $Y \succ_S X$.

# Integrating Prospect Theory ...

**Step 1: Combine Absolute & Contrast Utilities.**

People don't care *solely* about gains and losses. Usable economic theory must have both.

Imagine breaking down the $u(c|r)$ function:

$$u(c|r) \equiv m(c) + n(c|r),$$

where $m(c)$ is "consumption utility", and

$n(c|r)$ is "gain-loss utility".

Suppose $m(c)$ has "classical" properties (e.g., differentiable, increasing, concave, quasi-concave)

Claim: when $m(c)$ is approximately linear; if (dimension by dimension) $n(c|r)$ meets properties A0-A4, then, fixing $r$, $v(c|r) = u(c|r) - u(r|r)$ will (dimension by dimension) meet A0-A4.

Assume that $n_k(c_k - r_k|r) \equiv \mu(m_k(c_k) - m_k(r_k))$, where $\mu(x)$ is a "universal gain-loss function" meeting properties A0-A4.

Note: $u(c|c) = m(c)$ under this construction.

Predicts reference-free theories of utility correct when people consume at their reference levels!

## Step 2: But what is the Reference Point?

The literature has historically fluctuated between imprecision and status quo. A typical approach to this issue is in Tversky and Kahneman (1991):

> "A treatment of reference-dependent choice raises two questions: what is the reference state, and how does it affect preferences? The present analysis focuses on the second question ... the determinants of the reference state lies beyond the scope of the present article. Although the reference state usually corresponds to the decision maker's current position, it can also be influenced by aspirations, expectations, norms, and social comparisons."

Unexpectedly losing $50, or a mug you expected to keep, feels like a loss. Gaining $50 unexpectedly or getting mug unexpectedly are gains.

*But* spending $50 you planned to spend does not feel like a loss. Losing a mug you planned to sell? Getting $50 allowance you get every week? Getting delivery of mug long anticipated?

Given preponderance of theory, interpretations, and evidence saying status quo is reference point, *how can we say not the status quo?*

Answer 1: Because much of evidence on surprises, where status quo = expectations.

Answer 2: In fact, often wording obscures things.

Answer 3: Monetary preferences surely variant of expectations-based preferences.

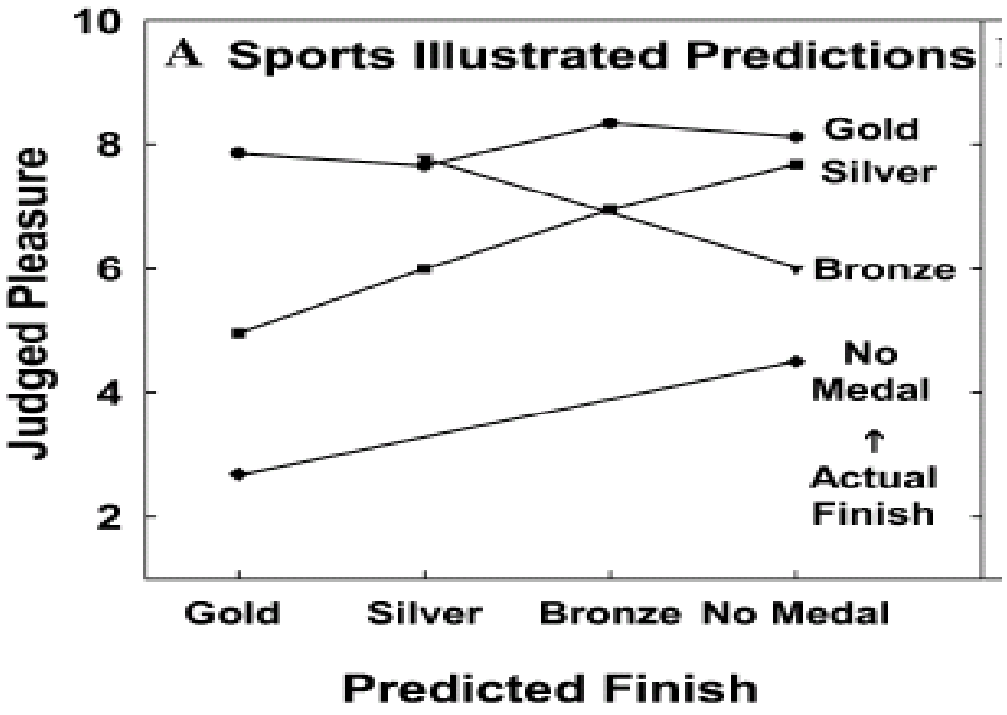
> Money is ... news about future consumption.

More intuition?

How feel about unexpectedly small wage raise?

Folk widsom: to avoid disappointment, don't get own (or your child's, or your friend's) hopes up.

McGraw, Mellers, and Tetlock (1995) replication, elaboration of famous Medvec et al experiment:





Role of expectations clear, but obscures punchline of earlier experiment: because expectations lower, average bronze medalist less happy than silver medalists!

# Risk Preferences Revisited

Koszegi & Rabin (2007): take these two steps.

Implications? Replicates "status quo" Prospect Theory in some *identified* contexts, classical DMU(W) in other *identified* contexts, and neither in yet other (*identified* ) contexts.

Large-stake risks → consistent with classical DMU(W)

Surprise, modest-stake risks → replicates "classical" Prospect Theory.

**But: expected**, modest-scale risks (even losses) → heavy risk aversion.

> Big issue in economic behavior: deductible aversion, extended warrantees, etc.

# Big Conceptual Issue

Our expectations depend on our plans ...

> Choice depends on preferences, and preferences depend on expectations of outcome. But (rational) expectations of outcome depend on actual choices.

xpectations of outcomes → preferences → choice → anticipated choice → expectations of outcome.

Yikes! Need some notion of consistent plans ... given preferences induced by plans, will you follow through on plans?

Punchlines of this approach:

1. Departures from classical predictions if and only if surprise or uncertainty.
2. (when uncertainty) environment/context seemingly influences preferences!
3. Some desirable plans not consistent ...

Illustrate:

## Shopping for shoes

Suppose value of shoes to consumer is $1.

Based on analysis using formalization of the psychology outlined above ...

Buying at price $p$ when fully expect to face price $p$ is the optimal consistent plan iff $p \leq 1$.

That is, Punchline 1: When no uncertainty, consumer WTP is "intrinsic" value for shoes.

But things change dramatically when uncertainty over prices. Suppose consumer knows:

$prob(p = p_H) = prob(p = p_L) = .5$.

If $(p_L, p_H) = (\frac{9}{8}, \frac{9}{8})$, Never Buy, $U = 0$.

If $(p_L, p_H) = (\frac{2}{8}, \frac{9}{8})$, Always Buy, $U = -\frac{1}{8}$

If $(p_L, p_H) = (0, \frac{9}{8})$, Buy iff $p = p_L$, $U = -\frac{1}{16}$

No uncertainty, don't pay more than intrinsic value.

Observe the environmental endogeneity: price distribution influences your WTP!

Utility can be lower buying than never buying ... and yet you buy! Consistency constraint in force!

$(p_L, p_H) = (\frac{7}{8}, \frac{7}{8}) \rightarrow$ always,

$(p_L, p_H) = (\frac{7}{8}, \infty) \rightarrow$ never

Though price is below intrinsic value, don't buy — because likely you wouldn't buy, less attached, and the idea of spending money more aversive to you.

**Another Punchline:**

Another punchline embedded in all these examples: Fundamental Principle: News feels bad on average.

Why? Because (under rational expectations) on average beliefs go up or down.

**But the big one is Punchline 2 from above:**

Because economic setting influences expectations, as in several strands of behavioral-economic research, preferences seemingly depend on context.

*Seemingly* because ... really 'stable' once defined more fully with arguments that economists used to ignore.

Demand for shoes? Not just price, but expected price matters.

Reinterpretation of the endowment-effect evidence.

The endowment effect is commonly explained by assuming that the reference point of ...

- ► ... ‘sellers’ is to keep the mug.
- ► ... ‘buyers’ is to keep their money.

Experimental protocol may (arguably) make natural expectations = endowments. But in real-world markets, expectations by those we *observe* as sellers and buyers *definitely* not endowments.

## Economics of Expectations and Reference Dependence

- Loss aversion in markets. Heidhues and Koszegi (2008, *AER*) show (among other things) how LA leads to "price stickiness" of various sorts in monopolistically competitive markets.
- Disposition effects. Genesove and Mayer on houses, and Odean on stocks: hold onto assets that have lost money.
- Small scale risk attitudes. Extended warrantees, deductibles on insurance, over-caution in purchasing uncertain items.
- Taxi cabs, etc. Camerer et al (1997) and others: "discretionary laborers" seem to work towards daily targets
- Principal-agent models. In various ways, won't fine tune incentives to *noisy* signals of effort as much as classical theories predict.
- Mas (2006), "Pay, Reference Points, and Police Performance": worker morale (how many tickets do you write!) effected by the *surprise* element in compensation—as identified by final-offer arbitration rulings.

■ Card and Dahl (NBER working paper, November 2009): Domestic violence increases when football team *unexpectedly* loses.

■ Precautionary savings and other consumption-savings implications: unambiguous, first-order precautionary savings ... rather than ambiguous (third derivative!), second-order of classical model.

■ News-utility perspective captures an old intuition: for small 'windfall', surprise wealth increase, consume it immediately. For surprise losses, decrease consumption later.

■ I'm sure half the conference was about the crisis ... me on the bandwagon: news-utility models predicts (especially rich people) were unhappy in September and October 2008?

## Monetary preferences

Prospect theory originally defined over money. But most formulations of reference-dependent utility defined over consumption. Most researchers (myself included!) cheat by going back and forth. But the news-utility perspective provides "foundations" for this cheat... Money as news: Changes in wealth are news about future consumption.

## Issues for empirical research

Since expectational environment matters, principled, theory-guided caution in when to extrapolate findings from one (campus or non-campus) situation to another (campus or non-campus) situation.

And since surpriseness of changes matters a lot, additional caution in using "natural experiments", very often surprises, to infer exogenous but non-surprise effects.

Directly measuring beliefs takes on additional value.

# Belief-Based Utility More Generally

The class of preferences receiving increased attention: belief-based preferences. People care about beliefs not solely for "instrumental" reasons.

Part of growing literature of preferences depending directly on (probabilistic) beliefs about the world,

Ego utility/self image/identity

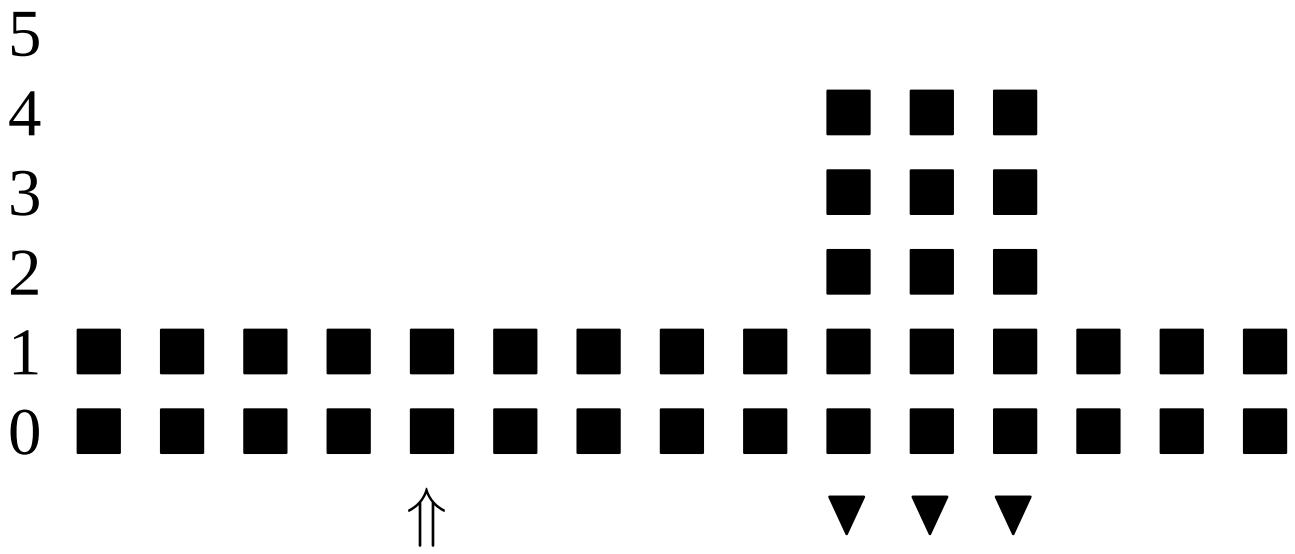
Anticipatory utility

Social comparison

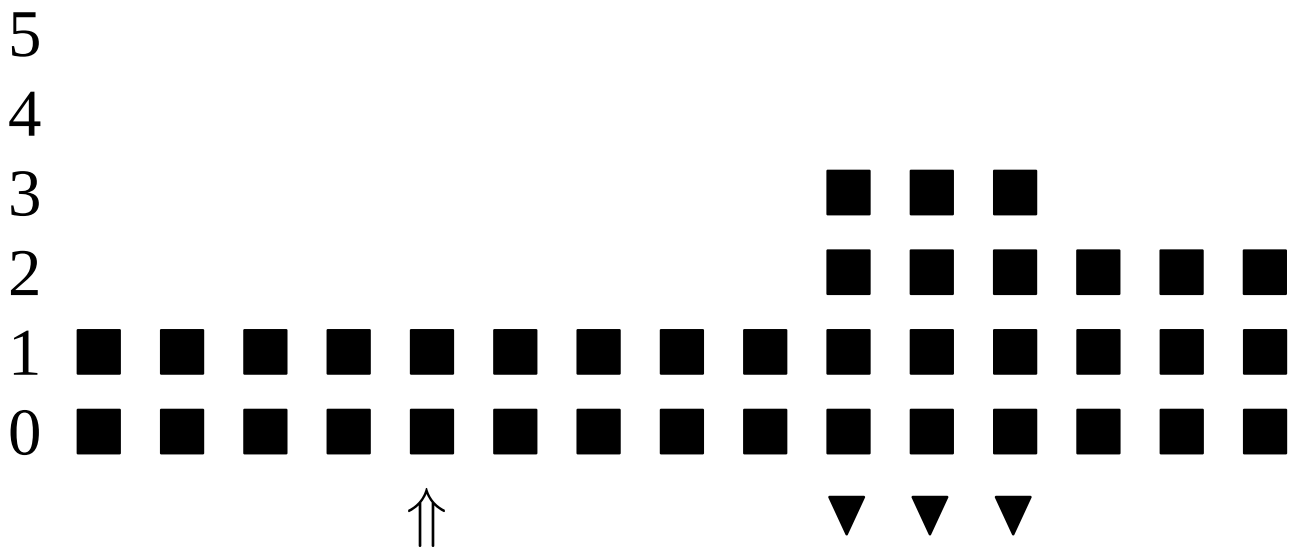
etc.

Let ⇑ be news of and ▼▼▼ duration of a vacation. Utility in time?

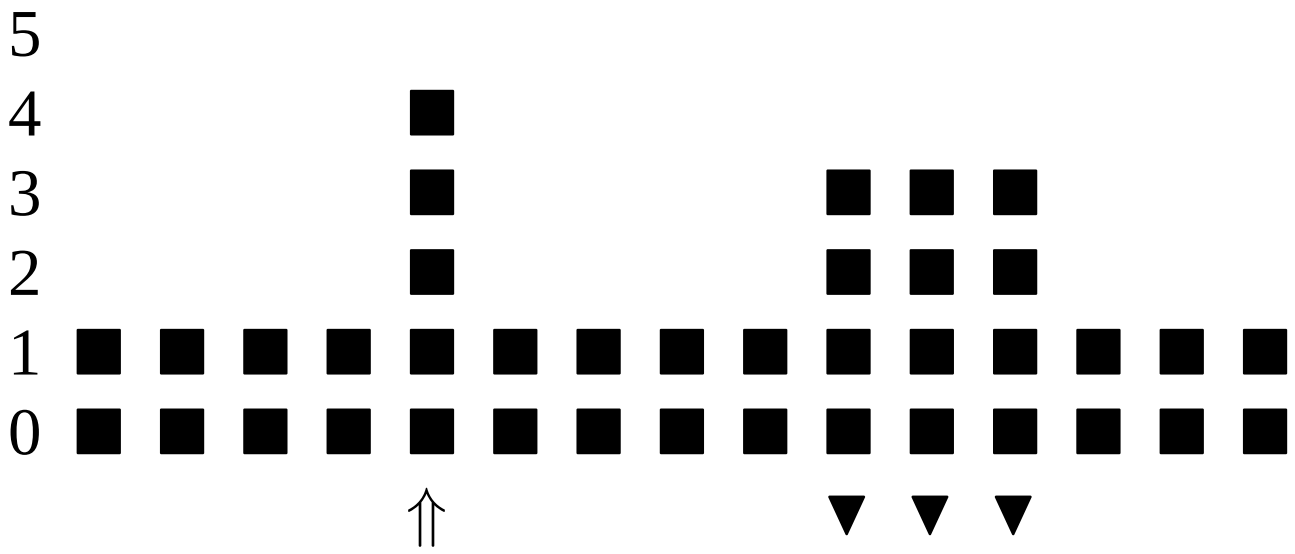
Could be:



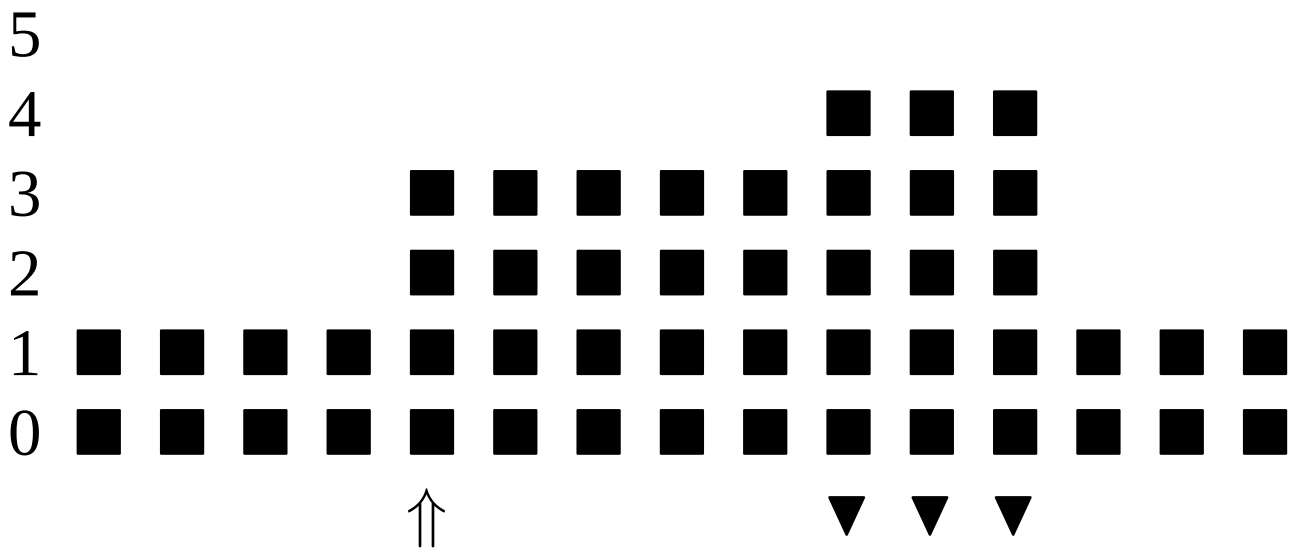

Or could be:




Or could be:




Or could be:



So what? Why do we care about the timing or reason for enjoying a vacation?

Often: we don't. All captured by $utility(vacation)$.

But suppose: "Club Medic" holiday package costs $10,000. And that: *Total* anticipatory utility plus consumption and remembered utility well worth it, but without anticipation *not* (nearly) worth it.

Case 1: No refunds.

Case 2: all but $50 refundable with 24 hours notice.

What do in Case 1? In Case 2?

Prediction under Case 2: Won't sign up!

# Introduction to Social Preferences

Moe: [from the bushes] Lisa!

Lisa: Moe?

Moe: Listen, I don't like you, you don't like me, but we both want to stop Homer from shooting a turkey.

Lisa: You don't like me? I like you.

Moe: You do? Then I like you, too. Here, have a towelette.

Among the most famous passages in economics is from The Wealth of Nations:

> It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard for their own interest. We address ourselves not to their humanity, but to their self-love, and never talk to them of our necessities, but of their advantage.

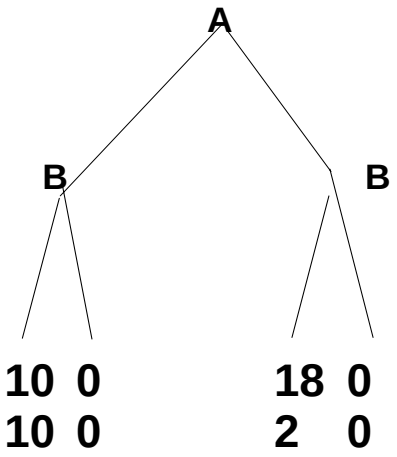
A sensibility for research recognizing the centrality of self-interest without ignoring the relevance of social preferences is illustrated by Dawes and Thaler (1988): Nailed-down cash boxes for roadside produce.

While moving now to the ‘field’, most evidence gathered from lab. I’ll report that. Legitimate concerns about extrapolability from one (campus *or* non-campus) setting to any other (campus or non-campus) setting.

And legitimate worries extrapolating one ‘scale’ to another.

Nonetheless: lots of evidence, and lots of intuition, about departures pure self interest.

# Two examples, & competing hypotheses.




| | C | D |
|---|---|---|
| C | $4,$4 | $0,$5 |
| D | $5,$0 | $1,$1 |

People (sacrifice money to) reject lop-sided offers in ultimatum game. Why?

*H1a* : failed, mischievious, or lazy selfishness.
*H2a* : punishing obnoxious/unfair behavior.
*H3a* : hate coming out behind random other subject.

People (sacrifice money to) cooperate with others in the one-shot prisoner's dilemma if and *only if* they think partner will. Why?

*H1b* : mistargeted attempt at repeated-game selfishness.
*H2b* : positive reciprocity, rewarding other.
*H3b* : implementing equitable outcome.

*H1a/b* : great pedigree, popularity, and wrongness.

*H2a*, *H2b*, *H3a*, and *H3b* all reasonable.

# Laboratory Social Preferences

Three classifications, “distributional preferences”, “intentions-based preferences,” and “other belief-based preferences”.

Note: even when abandoning self interest, economists prone towards particular assumption: altruism, either targeted (children) or untargeted (charity).

In all examples, proportions of laboratory subjects choosing each of two payoff combinations in terms of money (usually in units of U.S. pennies or Spanish pesatas) are drawn underneath the amounts.

From Kritikos & Bolle (2001), Charness & Grosskopf (2001), and Charness & Rabin (2002,2005):

# Distributional Preferences

Distributional preferences are roughly when we can represent Person 0's preferences by $U_0(\pi_0, \pi_1, \pi_2, ...)$, where $\pi_k$ is k's "material utility/payoffs". Begin with:

"Disinterested" distributional preferences"

Formally, $W_0(\pi_1, \pi_2, ...)$. Some examples of possible (extreme) preferences:

Surplus-maximizing: $W_0 = \sum_k \pi_k$.

Rawlsian/maximin: $W_0 = Min\{\pi_k\}_k$

Egalitarian: $W_0 = -\sum_k (\pi_k - \overline{\pi})^2$

Virtually no evidence on this! Only I know of:

| | | | |
|---|---|---|---|
| C chooses (A,B) allocation of | (400,400) | vs. | (750,375) |
| C-R | 46% | | 54% |
| C chooses (A,B) allocation of | (400,400) | vs. | (1200,0) |
| C-R | 82% | | 18% |

People seemingly strongly "Rawlsian".

"Non-disinterested" distributional preferences:

$$U_0(\pi_0, \pi_1, \pi_2, ...) =$$
$$(1-k-l)\pi_0$$
$$+k \cdot W_0(\pi_0, \pi_1, \pi_2, ...)$$
$$+l \cdot D_0(\pi_0 - \pi_1, \pi_0 - \pi_2, ...),$$

where $k, l, k+l \in [0,1]$.

$(1-k-l)\pi_0$ is self-interest

$k \cdot W_0(\pi_0, \pi_1, \pi_2, ...)$ is disinterested principles

$l \cdot D_0(\pi_0 - \pi_1, \pi_0 - \pi_2, ...)$ is "social comparison"

Fehr & Schmidt (1999), Charness & Rabin's (2002) simplistic two-person preferences:

$U_B(\pi_A, \pi_B) \equiv \rho\pi_A + (1-\rho)\pi_B$ when $\pi_B \geq \pi_A$.

$U_B(\pi_A, \pi_B) \equiv \sigma\pi_A + (1-\sigma)\pi_B$ when $\pi_B \leq \pi_A$.

Surely universal: $\rho \geq \sigma$. Possible variants:

$\rho = \sigma = 0$

$\rho = 1, \sigma = 0$ or $\rho = \sigma = \frac{1}{2}$ or $\rho = -\sigma = \infty$.

$1 \geq \rho \geq \sigma \geq 0$

$1 \geq \rho \geq 0 \geq \sigma$

$0 \geq \rho \geq \sigma$

# Evidence on Distributional Preferences?

Quick, selective examples meant to illustrate.

To ask purely distributional preferences, free of reciprocity, consider first only "dictator" games.

Evidence on $\rho$?

| | | | |
|---|---|---|---|
| B chooses (A,B) allocation of | (200,700) | vs. | (600,600) |
| C-R | 27% | | 73% |
| B chooses (A,B) allocation of | (0,800) | vs. | (400,400) |
| C-R | 78% | | 22% |

Crude summary of accumulated evidence: median $\rho$ about .4; 10% of subjects $\rho < 0$.

# Evidence on $\sigma$?

Warning: some of these data might be misleading, compared to other experiments, relatively little “Pareto damage”.

## Small sacrifice to avoid coming out behind?

| | | | | |
|---|---|---|---|---|
| B chooses (A,B) allocation of | (1200,625) | vs. | (600,600) |
| C–G | 88% | | 12% |
| B chooses (A,B) allocation of | (750,400) | vs. | (375,375) |
| C-R | 77% | | 23% |

## Significant sacrifice to avoid coming out behind?

| | | | |
|---|---|---|---|
| B chooses (A,B) allocation of | (4,1) | vs. | (0,0) |
| K-B | 88% | | 12% |

| | | | |
|---|---|---|---|
| B chooses (A,B) allocation of | (3,2) | vs. | (0,0) |
| K-B | 100% | | 0% |

| | | | |
|---|---|---|---|
| B chooses (A,B) allocation of | (800,200) | vs. | (0,0) |
| C-R | 100% | | 0% |

Small sacrifice to come out *further* behind?

| | | | |
|---|---|---|---|
| B chooses (A,B) allocation of | (625,625) | vs. | (1200,600) |
| C-G | 33% | | 67% |
| B chooses (A,B) allocation of | (400,400) | vs. | (750,375) |
| C-R | 55% | | 45% |

Recall: 46% of *disinterested* choose (400,400) ...

*Costlessly* take $ from other to avoid behind?

| | | | |
|---|---|---|---|
| B chooses (A,B) allocation of | ($X$,0) | vs. | (0,0) |
| K-B | 75% | | 25% |
| B chooses (A,B) allocation of | (900,600) | vs. | (600,600) |
| C-G | 67% | | 33% |
| B chooses (A,B) allocation of | (750,400) | vs. | (400,400) |
| C-R | 68% | | 32% |
| B chooses (A,B) allocation of | (2000,400) | vs. | (400,400) |
| C-R | 82% | | 18% |

Crude summary accumulated experimental evidence:

About 30% $\sigma < 0$, about 70% $\sigma > 0$.

Median $\overline{\sigma} > 0$, but very few $|\sigma| >> 0$.

# Intentions-Based Preferences

(By a long shot) not all social motivations are captured by distributional preferences.

People may care not *just* about outcomes, but with rewarding and punishing good and bad behavior.

Begin with the dark side: What induces "Pareto-damaging" behavior?

| A chooses | or | lets B choose | (800,200) | vs. | (0,0) |
|---|---|---|---|---|---|
| No choice | | | 100% | | 0% |
| fairer than | (800,200) | | 81-92% | | 8-19% |

| A chooses | or let | B choose | (750,400) | vs. | (375,375) |
|---|---|---|---|---|---|
| No choice | | | 77% | | 23% |
| fairer than | (750,400) | | 71% | | 29% |
| (400,750) | | | 80% | | 20% |

When free, how respond to goodness, badness?

| A chooses or lets B choose | (750,400) | vs. | (400,400) |
|---|---|---|---|
| No choice | ≈ 60% | | ≈ 40% |
| (550,550) | ≈ 55% | | ≈ 45% |
| (750,0) | 94% | | 6% |

What is going on? 1st vs. 2nd row? 1st vs. 3rd?

Crude summary of accumulated experimental evidence: Not much Pareto-damage without reciprocity; hint of increase in Pareto-damage by B if A is mean/unfa stronger indication of diminishing Pareto-damage if A behaves nicely.

How does good and bad behavior by one player affect the other player's inclination to engage in helpful sacrifice?

| A chooses | or | lets B choose | (750,375) | vs. | (400,400) |
|---|---|---|---|---|---|
| No choice | | | 46% | | 54% |
| (750$\pm$50,0) | | | 37% | | 63% |
| (550,550) | | | 11% | | 89% |

Comment on 1st vs. 2nd line: Whoa!

Crude impressions of accumulated evidence: Lots of helpful sacrifice. *Not* increased by other's good behavior, but withdrawn if other misbehaves.

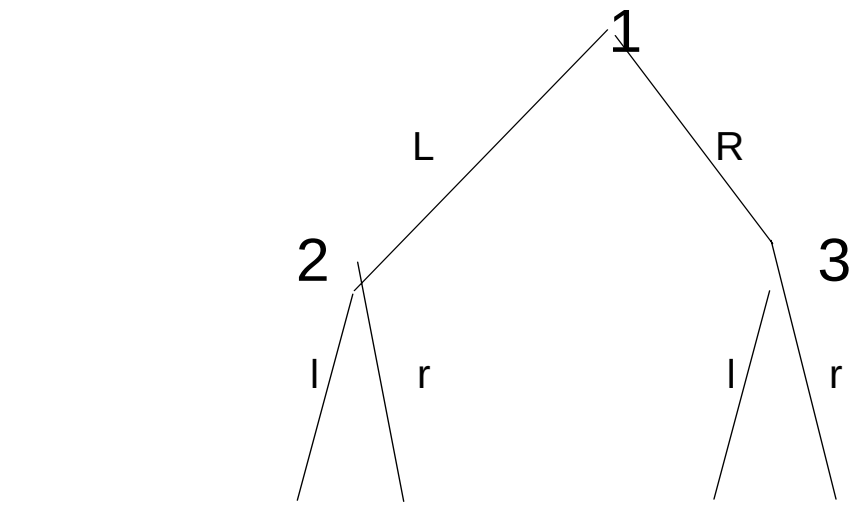
Approximately:

In UG: All negative reciprocity, no behindness aversion.

In PD: All equity, no positive reciprocity.

# Modeling Reciprocity? Hard.

Claim: Often no $U_j(\pi_i, \pi_j)$ can capture preferences.




What does Player 3 do if Player 1 goes R?

If Player 3 thinks Player 2 would have gone L?

If Player 3 thinks Player 2 would have gone R?

I claim there is something not quite right about those questions or answers.

## Self-Signaling and Other Signaling

Even when no reciprocity, social preferences not simply $U_j(\pi_i, \pi_j)$. Something else going on.

Lovely experiment by Dana, Weber, and Kuang (2007):

(other, self) of $A \equiv (\$5, \$5)$ vs. $B \equiv (\$1, \$6)$.

26% choose $B \equiv (\$1, \$6)$.

(other, self) of $C \equiv (\$1, \$5)$ vs. $D \equiv (\$5, \$6)$.

100% choose $D \equiv (\$5, \$6)$.

Now: people told to choose either \$5 or \$6 for self. 50% chance it is $A \; vs. \; B$, 50% $C \; vs. \; D$.

Subjects could choose without knowing which, or could for *free* reveal other's payoff first.

**37% choose to Not Reveal, take \$6.**

7% choose to Not Reveal, take \$5.
56% choose to Reveal.
If revealed and saw (\$5,\$5) vs. (\$1,\$6),
25% choose (\$1, \$6)
If revealed and saw (\$1,\$5) vs. (\$5,\$6),
90% choose (\$5, \$6).

% Choosing B over A goes 26% to 63%!

Interpretation? Cannot explain by simple distributional preferences. Not revealing payoffs is “moral wiggle room” to preserve self-image.

The Social Preferences literature marches forwards ...

Atlanta3.tex

January, 2010

# "Quasi Utility Maximization"

1. Focusing Effects and Narrow Bracketing

2. Misprediction Preferences

Behavior reflects maximization of coherent (and true!) $U(\cdot)$ at each moment in time.

> First example: Whenever focusing on particular choice, we maximize $U(\cdot)$, but we don't combine our separate choices or even pay attention to all possible choices.
>
> Second example: Even when our preferences will predictably change, we maximize PDV of current preferences.

# Focusing Effects, Narrow Bracketing

Life's a Bitch (of a complicated expected-utility maximization problem),

and then (millions of isolated decisions taken and billions of potential decisions untaken later)

you die.

Life is an infinite series of (potential) choice sets, $X_1$, $X_2$, ... $X_N$, ...When facing $X$ and $Y$ :

**Should:**

Max $_{x,y \in X \times Y}$ $u(x, y)$.

**Instead might:**

"Choose" some $\overline{x} \in X$ without thinking,

**Or:**

Max $_{x \in X}$ $u(x)$, Max $_{y \in Y}$ $u(y)$

1. Focusing effects: We only consciously make choices in an infinitesimal percentage of the infinite number of choice sets we actually have available, and

2. Narrow bracketing: We don't fully integrate our decisions with other decisions even when clear our utility would be higher if we did.

Review of the preferences we've examined ...

Suppose surprised you with choice: 50/50 lose $80, or lose $35 for sure.

Per prospect theory, most people take gamble.

Suppose offered choice (15 apples, 0 apples) vs. (9 apples, 4 apples) for (self,anonymous other).

Per altruism, fairness, non-gluttony, most people choose (9,4).

Fine. Real preferences, purposeful behavior. De Gustibus ... And yet:

## Focusing Effects

**Choose 50/50 lose $80 over lose $35 for sure?**

Coins in your pocket? Could all take the sure loss $35, then play 50/50 lose $\pm\$40$ with the person next to you ... generates 50/50 lose $75, gain $5.

Inconsistent with rational, broad bracketing of *any* (non-news) preferences to behave risk-lovingly.

**(9 apples, 4 apples) over (15 apples, 0 apples)?**

Why not take 15 apples, and share 5 of them with another random stranger...

Basically, sacrificing both self interest and total surplus is inconsistent with rational maximizing *any* "broad" social preferences.

## Narrow Bracketing

Two general approaches to showing that people "narrowly bracket":

1. Direct—we show that people don't combine problems they'd be better off combining.

2. Indirect—combine presumptive facts about "background noise" to argue calibrationally that observed choices are "too non-linear" to be consistent with any "global utility function".

Note: One one hand, skipping over all discussion of general "framing effects". On the other hand:

# “The mother of all framing effects”

Tversky and Kahneman (1986): “Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.”

Choose between:

A: $240
B: (.25 +$1,000, .75 $0)

Choose between:

C: -$750
D: (.75 -$1,000, .25 $0.)

Behavior: 84% A over B, 87% D over C.

Claim: This is narrow bracketing!

Proof 1: same as subjects choosing only 1 pair.

Proof 2: distribution combined choices: 73% AD, 11% AC, 14% BD, 3% BC.

But:

AD is .75 -$760, .25 +$240.
BC is .75 -$750, .25 +$250.

Rabin and Weizsacker (2009) replicate, with real stakes paid off 100% of time.

Also present *the exact same* choices "broadly".

"Choose one of the following four":

AC: sure loss of £5.10
AD: 75% chance lose £7.60, 25% chance gain £2.40.
BC: 75% chance lose £7.50, 25% chance gain £2.50.
BD: a 56.25% chance to lose £10.00, a 37.5% chance to gain/lose £0.00, and a 6.25% chance to gain £10.00

Results:

| presentation: | separate | broad |
|---|---|---|
| | % subjects choosing | |
| $A$ and $C$ | 21% | 11% |
| $A$ and $D$ | 28% | 0% |
| $B$ and $C$ | 11% | 38% |
| $B$ and $D$ | 40% | 51% |

# So what?

People don't do something impossible ... completely integrate life choices! And we can trick people with just-so choice combinations.

Answer 1: yes, impossible. And so we study it.

Answer 2: in your face, and still don't integrate.

Answer 3: Not about these preferences, or this pair! Simple dominance with almost any preferences:

**Theorem**: Suppose that utility $v$ is not CARA (i.e., not $\beta - \alpha e^{-rx}$ for any $\beta, \alpha, r$). Then there is a pair of choices between binary 50/50 lottery and sure thing where the decision maker violates FOSD.

Intuition is quite simple, and does not depend on risk-lovingness. E.g. suppose

CE of 50/50 ($30,$70) ≈ $48 → pay $2 to avoid risk
CE of 50/50 ($0,$40) ≈ $14 → pay $6 to avoid risk.

So suppose simultaneously choosing between:

Choose between:

A: $15
B: 50/50 ($0,$40)

Choose between:

C: $47
D: 50/50 ($30,$70)

Person will choose AD →50/50 ($45,$85)

But BC → 50/50 ($47,$87).

**Remark:** violation dominance if *any* two choices have the relevant feature, and total $ and certainly total U will add up.

# "Indirect" Evidence

Even if don't observe choices that are going uncombined ...

Any plausible background risk often implies that any "global preferences" ought be locally linear ...

An underestimate of variability of stock portfolio:

98% of time $\Delta \notin [-.01\%, +.01\%]$ (uniform)
86% of time $\Delta \notin [-.1\%, +.1\%]$ (uniform)
17% of time $\Delta \notin [-1\%, +1\%]$ (bellish).

Bracketing calibrations ... If have 2:1 loss aversion over total changes in wealth *today* ... what should be reaction to individual bets?

If \$10,000 in stock market ...

should accept 50/50 $g$ \$10.25/ $l$ \$9.75 bet.

If \$100,000 in stock market...

should accept 50/50 $g$ \$103/ $l$ \$98 bet or
50/50 $g$ \$10.02/ $l$ \$9.98 bet.

And in social preferences:

Recall the 45% of subjects who choose (\$4, \$4) over (\$7.50, \$3.75) for random, anonymous others. Maximizing what? Claim: *Not* "global Rawlsian" $Min[W_A, W_B]$, where $W_i$ is the Person i's wealth.

> Proof: If A is at least 25 cents poorer than B, then (\$7.50, \$3.75) better for $Min[W_A, W_B]$! And if at least \$4 poorer, (\$7.50, \$3.75) is much better. And at most 25 cents worse. But of course: essentially 50% A poorer by at least \$4. So $(7.50, 3.75)$ certainly better for $Min[W_A, W_B]$.

In fact, probably any $U(W_A, W_B)$ implies

$Max\ \Delta W_A + \Delta W_B$ in individual choices.

# Conclusions on Narrow Bracketing

"Classical" preferences could be maximized with much *easier* strategies.

If DMU(W) ... just be risk neutral over small stakes!

If selfish ... just be selfish!

If altruistic, maximize surplus!

People are (rationally or not) pursuing other goals.

Could be largely rational news-utility preferences.

But one way or another: modest-scale *narrowly bracketed* risk aversion a major fact about economic behavior, and major challenge to any attempt to model people's global maximization.

# Mispredicting Preferences

## Underappreciating changes in future preferences

Two ways tastes change over time:

1. Temporary Fluctuations (cues, satiation and deprivation, random time, moods, etc.)
2. Adaptation (paraplegia, standard of living), addiction, virtually any life event (good or bad) we (*un*)fully adapt to.

Changing tastes are fact about utility, not indication of irrationality.

The error: people systematically underappreciate (even very predictable) changes in their tastes.

# Fluctuations

Nothing in economics says that taste for food, water, sleep, or sex (the big four!) won't fluctuate greatly based on satiation and circumstances.

Yet, despite *vast experience* with these fluctuations, evidence suggests we underestimate these fluctuations.

E.g. when currently hungry, underappreciate how full we will feel after eating; when sated, underestimate hunger's influence on craving.

Evidence? Two beautiful experiments ...

Classic: muffins at the supermarket...

Read and van Leeuwen (1998): Office workers asked to choose between healthy snacks and unhealthy snacks that they would receive in one week, either at a time when they should expect to be hungry (late in the afternoon) or satiated (immediately after lunch).

% of Subjects Choosing “Unhealthy” Snack

| | | Future Hunger | |
|---|---|---|---|
| | | Hungry | Satiated |
| Current Hunger | Hungry | 78% | 56% |
| | Satiated | 42% | 26% |

Badger *et al* : Elicited (real stakes) WTP for *second* dose of heroin substitute BUP—reduces craving for heroin, aiding withdrawal—from long-time heroin addicts regularly receiving single dose.

Half asked when "deprived" about 2nd dose, half asked when "satiated" about 2nd dose.

Deprived: 2 hours before scheduled dose.

Satiated: right after scheduled dose.

Half asked willingness to pay for 2nd dose today, half about 2nd dose on next visit.

Average revealed value of a 2nd dose (always delivered in satiated state):

| | | When they would get the dose | |
|---|---|---|---|
| | | Today | Next visit |
| Current | Deprived | $75 | $60 |
| Craving | Satiated | $50 | $35 |

These WTPs are for the *exact same circumstances by highly experienced* addicts.

# Aside: Revealed Preference?

Which is the “true” value?

By BE principles ... probably the $35. *Not* because drugs are bad and this is lowest ...but because

a) present bias (next topic!) says 2nd column better than first, and b) projection bias says bottom row better than top.

Deciding what I want when in same emotional/craving state as the situation, but ahead of time so self-control (next topic!) not an issue, is often the ideal.

# Longer-Term Changes

Big-time major fact about people: We adapt to changes.

Another big-time (harder-to-prove, hypothesized) fact about people: We underestimate adaptation.

Classic study by Brickman, Coates, and Janoff-Bulman (1978) suggests that people underappreciate adaptation- winning lotteries, losing limbs.

Evidence on this is very hard.

# Behavioral economists love mugs ...

It's like heroin or borrelnoten to us!

Loewenstein and Adler (1995) randomly assigned subjects. A "prediction" treatment group were shown an embossed coffee mug and then told that they would later be given one as a prize but would have the opportunity to exchange it for cash.

| Condition | Prediction of Valuation | Actual Valuation |
|---|---|---|
| Prediction | $3.45 | $4.89 |
| No Prediction | ——— | $5.56 |

Interpretation: mini-look at adaptation. People underappreciate the endowment effect—how ownership will affect their valuation.

# Modeling Mispredicting Preferences

True preferences are given by:

$U^t = \sum_{\tau=t}^{T} \delta^\tau u(\mathbf{c}_\tau, \mathbf{s}_\tau),$

Let $\tilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau | \mathbf{s}_t) \equiv$ prediction time $t$ in state $\mathbf{s}_t$ of $u(\mathbf{c}_\tau, \mathbf{s}_\tau)$.

Rational expectations: $\tilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau | \mathbf{s}_t) = u(\mathbf{c}_\tau, \mathbf{s}_\tau)$.

Loewenstein, O'Donoghue, and Rabin (2002) assume that a person with *projection bias* understands the qualitative nature of changes in her preferences but underestimates the magnitude.

Definition: Predicted utility exhibits *simple projection bias* if there exists $\alpha \in [0, 1]$ such that for all $\mathbf{c}$, $\mathbf{s}$, and $\mathbf{s}'$, $\tilde{u}(\mathbf{c}, \mathbf{s} | \mathbf{s}') = (1 - \alpha)\, u(\mathbf{c}, \mathbf{s}) + \alpha\, u(\mathbf{c}, \mathbf{s}')$.

Then assume per usual: each $t$, person maximizes

$\widetilde{U}^t = \sum_{\tau=t}^{T} \delta^\tau \tilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau | \mathbf{s}_t).$

This slide is intentionally blank

# Errors because of such mispredictions?

## Addiction/Habit Formation

People may underappreciate the power of addiction. Under-investment in good addictions (healthy food, exercise, coffee) and over-investment in bad addictions (tobacco, alcohol)

Underappreciation of negative internalities:

Non-addict underestimates how bad she'll feel once addicted.
Addict underestimates how good she'll feel once unaddicted.

Underappreciation of habit formation:

Non-addicts underestimates how painful it will be to quit.
Addicts exaggerate how long the pain of quitting will last.

Underappreciating fluctuations, cues can also lead to binges and (costly) failed episodes of quitting.

Levy (2008) estimates role projection bias (and self-control problems) in U.S. tobacco consumption, and finds large role of misprediction: reaction by current non-addicts to permanent price changes indicates they don't think they'll become addicted by consuming now.

## Many Hard-to-Reverse Decisions

In a variety of situations, people make difficult-to-reverse decisions when they are in a “hot” state that is unlikely to persist.

Central intuition: don’t implement based on average feeling, but on the extreme feeling.

Getting married in the heat of passion
Committing suicide in the depth of depression
Sending that e-mail under the grip of rage

Purchase of a Durable Good

Projection bias leads a person to underestimate how quickly the ownership excitement will decay. Hence:

Too likely to buy the good.

## Benefits of Cooling Off

[Homer is trying to buy a handgun]

Cashier: Sorry, the law requires a five-day waiting period. We've got to run a background check.

Homer: Five days? But I'm mad now! I'd kill you if I had my gun.

Cashier: Yeah, well, you don't.

But he will in 5 days ...

Both normative policy analysis and descriptive political economy: cooling-off periods are widely used, and many advocate using them even more broadly.

Mandatory "cooling-off period" for sales contracts:

Generalization: we may often want to require "persistent choice": requiring people to say repeatedly over time that they want to a choice (or letting them back out) is superior to either bans or enforcing all market contracts. Marriage, suicide, durable goods, jobs, tattoos, etc.

## Excess wealth-seeking/consumption?

We don’t fully appreciate how much our pleasure from future standard of living will decrease once we become accustomed to that standard of living.

Recently, scale of economic growth as source of well-being (in already-wealthy nations) called into question. Misprediction might play a role.

# Close Cousins to this Error

Projection bias is a failure of contingent thinking, in particular inability to take self out of current state.

More famous cousin: "information projection" in cognition:

Intrapersonal empathy gaps may lead to inter-personal empathy gaps.

But converse: others who are/have been in situation may be better at predicting our tastes in same situation than we are.

Atlanta4.tex

January, 2010

# "Present-Biased Preferences"

(What David calls hyperbolic or quasi-hyperbolic discounting)

Moe: "I just bought a deep fryer that can deep fry a water buffalo in 40 seconds."

Homer: "40 seconds??? But I'm hungry *now*!"

Anonymous teacher evaluation: "The problem sets should have been graded. I had no incentive to do them, and as a result did poorly on the exams."

# The time-inconsistent taste for immediate gratification.

We tend to overpursue immediate gratification — relative to our *own* long-term preferences.

This is surely the most active and fruitful area of 3rd-wave behavioral economics.

Economists assume discounting is time-consistent: people have same preferences now over future behavior as we will have in the future. This is implied by our assumption of exponential discounting.

*Not* a quote from Wealth of Nations: "It is not for current pleasure that the baker is drunk and the brewer is fat, but for their love of their long-run well-being ..."

Exponential model says *a priori* : people smokers, overweight, in $9,500 of credit-card debt., earn less because never finishing high school, crack addicts, etc. only when expected pleasure they derive from such life courses exceeds expected costs.

Exponential model is wrong. So what? All models are! But: it is wrong in a systematic direction, leading to immensely misleading behavioral and welfare conclusions about immensely important economic topics.

And even small departure can have huge impact.

Instead: Present bias.

> We care more about today than the future. And today we care roughly equally about well-being on any two future dates, but when the future arrives and the first of these dates becomes “today”, then we care more about that first date than about the second date.

> We smoke more, eat more, and work less tomorrow than today we wish our tomorrow selves to do.

# Modeling Present-Biased Preferences

Simple Model (Laibson, 1994): $u_t$ is a person's well-being on Day $t$. $U^t$ is the person's intertemporal preferences from the perspective of period $t$.

$$U^t(u_t, u_{t+1}, ..., u_T) \equiv \delta^t u_t \ + \ \beta \sum_{\tau=t+1}^{T} \delta^\tau u_\tau.$$

where $\delta \approx 1$ is the standard discount factor.

How are these preferences time-inconsistent? Person cares $\frac{1}{\delta} \approx 1$ more about $u_\tau$ vs. $u_{\tau+1}$ if asked on Day $t < \tau$, but cares $\frac{1}{\beta\delta} > \frac{1}{\delta}$ more about $u_\tau$ if asked on Day $\tau$.

Why $\delta \approx 1$?

A list of arithmetic truths ...

$.99^{365\times 2} \approx \frac{1}{1536}$ $.99^{52\times 2} \approx \frac{1}{2.8}$

$.999^{365\times 2} \approx \frac{1}{2.1}$ $.999^{365\times 24\times 2} \approx \frac{1}{40{,}987{,}013}$

$.9999^{365\times 24\times 2} \approx \frac{1}{5.7}$ $.999999^{365\times 24\times 90\times 2} \approx \frac{1}{5}$

How much favor well-being *now* over well-being

a week from now?
tomorrow?
an hour from now?
40 seconds from now?

How much favor your well being a day 17 years from now over a day 19 years from now?

Even simpler calibration theorem! Most people have taste for immediate gratification, but *nobody* a non-negligible "taste for early future gratification".

Some essentially equivalent (true) statements:

Exponential wildly miscalibrated for short term.

People bothered *only* by delays in *immediate* gratification.

"Time-inconsistent taste" is redundant!

Evidence near-term impatience is evidence of present bias.

E.g., in three days when $\beta = \frac{9}{10}$ and $\delta = 1$:

$$U^1(u_1, u_2, u_3) = u_1 + \frac{9}{10}(u_2 + u_3)$$

$$U^2(u_2, u_3) = u_2 + \frac{9}{10}u_3$$

Day 1: Person cares equally about well-being on Day 2 and Day 3

But on Day 2, cares 11% more about well-being on Day 2 as on Day 3.

Leads to "preference reversals" and time-inconsistent behavior.

# Extended Example: Procrastination

Suppose with 120 minutes effort today could reduce by 10 minutes effort needed to undertake a task every day for rest of life.

Within 2 weeks, you will on net save time.
In a year, 58 hours. In a decade, 600 hours.

Suppose disutility of effort is (same) proportion to time each day.

No deadlines, no commitment devices.

Do you do the task? If so, when?

If do task today, your lifetime utility is:

$$U^t = -120 + \beta\delta \cdot 10 + \beta\delta^2 \cdot 10 + \beta\delta^3 \cdot 10 + ...$$

$$= -120 + \beta\tfrac{\delta}{1-\delta}10,$$

relative to the utility you would get from not doing.

Consider first exponential discounting.
$\beta = 1, \delta = .999$.

Then, utilities (relative to never doing the fix) are:

$$U^t(fix\ today) = -120 + \frac{.999}{1-.999}10 = 9,870.$$

$$U^t(fix\ tomorrow) = .999(-120 + \frac{.999}{1-.999}10) \approx 9,861$$

$$U^t(fix\ next\ day) = .999^2(-120 + \frac{.999}{1-.999}10) \approx 9,852$$

...

$$U^t(never) = 0$$

Person will do it right away.

Suppose some taste for immediate gratification:

$\beta = .9$, $\delta = .999$. Then:

$$U^t(fix\ today) = -120 + .9[\tfrac{.999}{1-.999}10] = 8,871$$

$$U^t(never) = 0$$

Even with taste for immediate gratification, worth doing. Choose Today *vs* Never $\rightarrow$ Do Today.

Even if it took 6 days, choose Now over Never!

> **LESSON 1: Present bias is <u>not</u> myopia: (We) people care a lot about the future; we just have moderate favoritism towards immediate vs. slightly delayed gratification.**

But:

$$U^t(fix\ tomorrow) =$$

$$.9 \cdot .999(-120 + \tfrac{.999}{1-.999}10) \approx 8,874 > 8871$$

You'd prefer to do tomorrow rather than today. Puts the -120 "inside the $\beta$".

Also note:

$$U^t(fix\ day\ after\ tomorrow) =$$

$$.9 \cdot .999^2(-120 + \tfrac{.999}{1-.999}10) \approx 8,865 < 8871.$$

So: preferences are:

$$U^t(fix\ tomorrow) \succ U^t(fix\ today) \succ$$
$$U^t(day\ after\ tomorrow) \succ U^t(never)$$

So: When do you do the task?

Answer: It depends.

If you think that not doing today means you will do tomorrow, then ... don't do today.

If you think not doing today means delaying two or more days, then *will* do today.

What you do depends on your beliefs about own future behavior.

What *should* you believe?

What *do* you believe?

# Awareness: Sophistication vs. Naivete

Two alternative assumptions about beliefs:

Sophistication: Fully aware of future self-control problems, so correctly predict how behave if wait.

Naivety: Naive about your future self-control problems, so believe you will behave in the future exactly as you now would like yourself to behave. That is, today $\beta < 1$, but think beginning tomorrow you'll have $\beta = 1$.

Partial Naivety: Important in many applications.

Back to example. Recall:

If $\beta = \frac{9}{10}$, then

$$U^t(fix\,tomorrow) \succ U^t(fix\,today) \succ U^t(never)$$

If $\beta = 1$, then

$$U^t(fix\,today) \succ U^t(fix\,tomorrow) \succ U^t(never)$$

So if naively think that tomorrow you will have a $\beta = 1,$ then you will not do today *believing* will do tomorrow.

But when tomorrow comes:

You will not do, planning to do the next day.

And when the next day comes:

You will not do, planning to do the day after ...

You will procrastinate forever—always planning to do the task the next day.

(Full) sophistication overcomes procrastination: you will never wait more than 2 days.

**LESSON 2: Some naivety about future self-control problems is realistic, and can matter a lot.** (But in lots of situations, matters very little; and sometimes, sophistication is worse!)

Procrastination forever, though by long shot wish to improve word-processing skills.

**LESSON 3: Observed 'Life Course,' if involves incremental decisions rather than once-and-for-all choice, may not reveal much about life-course preferences.** Lifelong tobacco and alcohol addiction, obesity, CC debt are not always fulfilment of life long goals.

Suppose now: benefit of 10 minutes is 7 days a week, but can only do fix at office on weekdays. On fridays, *can't* naively believe will do tomorrow! Waiting to monday not worth it ... will do on first friday!

**LESSON 4: Details of short-run incentives, not just the long-run, can matter a lot.**

Related: Estimating with misspecified $\beta = 1$ model, we'd interpret somebody 'unwilling' to spend 2 hours to save 10 minutes forever as extremely impatient. We'd think $-120 + \frac{\delta}{1-\delta} 10 \leq 0$, so never do task $only$ if $\delta \leq \frac{12}{13} \Rightarrow \delta^{365} \leq .000000000002$. If that is what we observed, we'd conclude insane impatience. By constrast, if we saw *same* person making a once-and-for-all choice to spend over 6 days to save 10 minutes a way, we'd figure out that $\delta^{365} > \frac{1}{2}$.

**LESSON 5: Because $\beta = 1$ is wildly miscalibrated to explain short-run impatience but fine for long-run, interpreting behavior through the misspecified $\beta = 1$ model will generate wildly inconsistent (and sometimes absurd) measures.**

The reason that it so often looks like people are impatient in the short run but patient in the long run is ... people are impatient in the short run but patient in the long run. This is a "paradox" in classical framework, but is essentially the definition of present bias.

Although massive mismatch between life-course behavior and life-course preference can happen even with sophistication, 'marker' for self control often when people do not expect behavior. People predict CC debt? Future weight? Smoking?

**LESSON 6: People may systematically and severely mispredict own future behavior.**

In many simple environments, "good behavior" based on *ratio* immediate costs to future benefits of good behavior, *not* on net benefits. So misbehavior not reduced "when matters". In fact, O'DR (2001) illustrate how procrastination can be worse for more important tasks! More important → intend greater effort to do right → procrastinate!

**LESSON 7: Though many mistakes likelihood of errors diminishes 'when it matters,' there is no strong presumption that self-control problems decrease for high stakes.**

# Many implications ... growing literature.

## Mechanism design in markets and government.

## **Default effects**.

Very simple logic: present-biased people may never switch one option to another, revealing almost nothing of their actual preferences. Welfare heavily affected by default when $\beta < 1$, but almost no effect when switching costs are low for $\beta = 1$.

## Incremental Incentives on longer term projects

Should give problem sets ... even with uncertainty, can be optimal to force interim progress.

## Deadline effects;

Like the "friday effect" from earlier, people do things at last minute, often at much higher costs, that impossible to explain with time-consistent preferences.

[Watching the local news on April 15, filmed live outside the springfield post office]

Homer: "Look at all those idiots who didn't file their taxes"
Lisa: "Did you do your taxes, Dad?"

Homer: "Yes. I did it over a year ago."

Lisa: "Dad—those were *last year*'s taxes! You have to file every year."

Homer: "D'oh!"