Introduction to Social Preferences⇌

- Social preferences can be defined as "ways that people's utility depend *directly* on the well-being, motives, and beliefs of others".⇌
- Among the most famous passages in economics is from Adam Smith's *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.

↬

Social Preferences

- Research recognizing the centrality of self-interest without ignoring social preferences, illustrated by Dawes and Thaler (1988):⇌
  - "In the rural areas around Ithaca it is common for farmers to put some fresh produce on the table by the road. There is a cash box on the table, and customers are expected to put money in the box in return for the vegetables they take.⇌ The box has just a small slit, so money can only be put in, not taken out.⇌ Also, the box is attached to the table, so no one can (easily) make off with the money.⇌ We think that the farmers have just about the right model of human nature.⇌ They feel that enough people will volunteer to pay for the fresh corn to make it worthwhile to put it out there.⇌ The farmers also know that if it were easy enough to take the money, someone would do so."

↬

Social Preferences

While moving now to the 'field', most evidence gathered from lab.$\rightleftharpoons$

- Legitimate concerns about extrapolability from one (campus *or* non-campus) setting to any other (campus or non-campus) setting.$\rightleftharpoons$

Note: Ignoring social preferences requires more than (wrong) claim that self interest **always** much more powerful. $\rightleftharpoons$

- Rather, *also* requires$\rightleftharpoons$
  - No situations where actors can significantly affect well-being of others with only small effect on own well-being.$\rightleftharpoons$
- Note: Reference Dependence massively important in social prefs$\rightleftharpoons$
  - We ignore.

$\looparrowright$

Social Preferences

I am attitudy and puzzled by much of the research.$\rightleftharpoons$

- A huge industry in experiments on social preferences.$\rightleftharpoons$
- Most active area of both experimental and behavioral economics.$\rightleftharpoons$
- But not the most successful.$\rightleftharpoons$
  - Little integration of any lab lessons into mainstream economics.$\rightleftharpoons$
  - Movement towards conceptual tightness disappointingly slow.$\rightleftharpoons$
  - Movement towards serious empirical conclusions (including acknowledging ones staring us in the face) distressingly slow.$\rightleftharpoons$
  - And 100% self-interest not as far off the mark as many other assumptions behavioral economists seek to modify, so benefits lower.

$\looparrowright$

Social Preferences

More attitude:⇌

- Scientifically sinful to not disentangle obvious confounds when they *are* obvious, and when they are easy to disentangle.⇌
- Multiple theories for the same data rather than creating data to narrow down theories has been egregious.⇌
  - Even when data needed is really really really *really* easily created.⇌
- Part of reason: people sense lab data missing something of people's intuition about the world.⇌
  - Can't get envy in lab...mimic it with other social preferences!⇌
  - But that suggests not to use the lab—it does not justify adding confounds to the lab so that the data match your intuition.

↬

Social Preferences

Two examples, & competing hypotheses.⇌

- Ultimatum game and Prisoner's Dilemma⇌

	Accept	Reject
Share	5,5	5,5
Grab	8,2	0,0

	C	D
C	4,4	0,5
D	5,0	1,1

⇌

- Variants of these two gamesdominated much of the earlier social-preferences literature ...⇌
  - Despite being awful games to do science of social preferences with.⇌
- Now we have Dictator and Trust games!

↬

Social Preferences

Consensus empirical fact: people often (sacrifice money to) reject lop-sided offers in ultimatum game.$\rightleftharpoons$Why?$\rightleftharpoons$

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

Consensus empirical fact: people often cooperate with others in one-shot prisoner's dilemma, but *only if* they think partner will. $\rightleftharpoons$Why?$\rightleftharpoons$

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

H1a/b: great pedigree, popularity, and wrongness.$\rightleftharpoons$
H2a, H2b, *H3a,* and *H3b* all reasonable. $\rightleftharpoons$

- And could all be true

↬

Social Preferences

Laboratory evidence has inspired many models of "social preferences".⇌

- Three classifications, "distributional preferences", "intentions-based preferences," and "other belief-based preferences".⇌
- Note: economists prone towards particular assumption: altruism, either targeted (children) or need-based (charity).⇌
  - Both are majorly right. Economists should and do study targeted altruism, and nobody objects to the investigation of charity, etc.⇌
- But lab research has emphasized fairness, reciprocity, etc.

↬

Social Preferences

Distributional Preferences$\rightleftharpoons$

- Represent Person 0's preferences by $U_0(\pi_0, \pi_1, \pi_2, ...)$, $\rightleftharpoons$
  - where π_k is Person k's "material" utility/payoffs. Begin with:$\rightleftharpoons$
- "Disinterested distributional preferences": Allocations people choose for others, when their choices do not affect their own outcome?

↬

Social Preferences

Formally, $W_0(\pi_1, \pi_2, ...)$. Examples of (extreme) preferences:$\rightleftharpoons$

- Surplus-maximizing: $W_0 = \sum_k \pi_k$.$\rightleftharpoons$
  - (If π_k really "material hedonic return", then this is utilitarianism.)$\rightleftharpoons$
- Rawlsian/maximin: $W_0 = Min\{\pi_k\}_k \rightleftharpoons$
- Egalitarian: $W_0 = -\sum_k (\pi_k - \overline{\pi})^2 \rightleftharpoons$

Note: $\rightleftharpoons$

- Rawlsian preferences have a form of aversion to inequality, but they are monotonically increasing.$\rightleftharpoons$
- "Egalitarian preferences" are qualitatively more extreme dislike of unequal outcomes; you'd actually lower people's payoffs.)$\rightleftharpoons$

These are of course unrealistically extreme forms.

$\curvearrowright$

Social Preferences

"Non-disinterested" distributional preferences:$\rightleftharpoons$ $U_0(\pi_0, \pi_1, \pi_2, ...) =$

$$(1 - k - l)\pi_0$$

$$+(k \cdot W_0(\pi_0, \pi_1, \pi_2, ...)$$

$$+(l \cdot D_0(\pi_0 - \pi_1, \pi_0 - \pi_2, ...),$$

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

- Of course, components are interpretation; don't observe separately. $\rightleftharpoons$
- What are they?$\rightleftharpoons$
- Answer: $\rightleftharpoons$ self interest, $\rightleftharpoons$ disinterested principles$\rightleftharpoons$ "social comparison"

↬

Social Preferences

CR's simplified parameterization of two-person preferences, based on FS.$\rightleftharpoons$

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

Exaggerates the kinkiness of preferences.$\rightleftharpoons$

- This is largely for simplification.$\rightleftharpoons$
- Surely generally true that $\rho \geq \sigma$.$\rightleftharpoons$
- But the rest up for empirical grabs.

$\looparrowright$

Social Preferences

A bunch of (not necessarily mutually-exclusive) examples:$\rightleftharpoons$

- $\rho = \sigma = 0 \implies$ pure self interest.$\rightleftharpoons$
- $\rho = 1, \sigma = 0 \implies$ disinterested pure Rawlsian$\rightleftharpoons$
- $\rho = \sigma = \frac{1}{2} \implies$ dinterested surplus maximizer$\rightleftharpoons$
- $1 \geq \rho \geq \sigma \geq 0 \implies$ Social-Welfare Preferences$\rightleftharpoons$
- $1 \geq \rho \geq 0 \geq \sigma \implies$ inequity averse$\rightleftharpoons$
- $\rho \geq 1 \geq 0 \geq \sigma \implies$ egalitarian?$\rightleftharpoons$
- $0 \geq \rho \geq \sigma \implies$ competitive?$\rightleftharpoons$

$\looparrowright$

Social Preferences

Lab Evidence on Social Preferences⇌

- In all examples, proportions choosing each of two money combinations (usually U.S. pennies or Spanish pesatas) are drawn underneath the amounts.⇌
  - From Kritikos & Bolle (2001), Charness & Grosskopf (2001), and Charness & Rabin (2002,2005):⇌
- Caution:⇌
  - These slides somewhat dated⇌
  - More evidence on some of this⇌
  - (Although remarkably little)

↬

Some Lab Data on Distributional Prefs

Evidence on disinterested distributional preferences?

- There is virtually no $-stakes evidence on disinterested preferences!
- From Charness and Rabin (2002):

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

Keep these in mind.

- Especially the first one
- Disinterested care a lot about "equity"

Social Preferences

Evidence on Non-Disinterested Distributional Preferences?

- Quick, selective examples meant to illustrate. $\rightleftharpoons$
- (But meant not to be misleading.)$\rightleftharpoons$
- To ask purely distributional preferences, free of reciprocity, consider first only "dictator" games.$\rightleftharpoons$
  - My attitude begins here ... $\rightleftharpoons$
  - disentangle rather than confound.

$\hookrightarrow$

Social Preferences

What is the evidence on ρ?

B chooses (A,B) allocation of	(200,700)	vs.	(600,600)
Old C-R:	.27		.73

B chooses (A,B) allocation of	(0,800)	vs.	(400,400)
Old C-R	.78		.22
New C-R	.56		.44
New C-R with requests by A	.45		.55

My impressions from the accumulated experimental evidence:

- The average or median ρ is about .4.
- About 10% of subjects have $\rho < 0$.

Social Preferences

Evidence on σ?$\rightleftharpoons$

- Shockingly (unforgivably) little evidence on this disentangled from negative reciprocity.$\rightleftharpoons$
  - Do people Pareto-damage when other has done nothing wrong?$\rightleftharpoons$
- Warning: might be misleading, because compared to other experiments, relatively little "Pareto damage".$\rightleftharpoons$

$\looparrowright$

Social Preferences

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%*	

That 23% is the largest % I know of for reciprocity-free strict Pareto-damaging sacrifice, across all experiments ever run.

- Recent evidence, still absorbing, finds more.⇌
- (And certainly also more evidence of less.)

↬

Social Preferences

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%*	

↬

Social Preferences

Small sacrifice to come out *further* behind?⇌

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

- - Recall 1: Only 23% choose (375,375) over (750,400).⇌
  - Recall 2: 46% of *disinterested* choose (400,400) over (750,375).⇌
- **So:** more people sacrifice to come out **behind** when it really helps other than will sacrifice to avoid coming out behind! ⇌
- **And**: most who do not sacrifice to help:⇌ are refusing out of sincere Rawlsian motives, not because selfish or because hate coming out behind.

↬

Social Preferences

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%*	⇌

- More than I ever expected... $\frac{1}{3}$ taking $ away from other.⇌
  - But recall: behindness aversion says motive super-strong here.

↬

Social Preferences

Crude summary accumulated experimental evidence:$\rightleftharpoons$

- About 30% $\sigma < 0$, about 70% $\sigma > 0$.$\rightleftharpoons$
- Median $\overline{\sigma} > 0$, but very few $|\sigma| >> 0$.$\rightleftharpoons$
- Including very few $\sigma << 0$.$\rightleftharpoons$

Update on original examples and hypotheses? So far:$\rightleftharpoons$

- Little indication that rejections in the ultimatum game have much to do with "behindness aversion". $\rightleftharpoons$
  - Maybe it is all about negative reciprocity instead. $\rightleftharpoons$
- But lots of eidence of sharing when ahead ... very consistent with PD being from "inequity aversion".$\rightleftharpoons$
  - Does not preclude positive reciprocity.

$\looparrowright$

Social Preferences

Intentions-Based Preferences⇌

- People may care not *just* about outcomes, but with rewarding and punishing good and bad behavior.⇌
  - In bilateral context, we think of this as reciprocity. ⇌
- I emphasize **"just"** to make clear: ⇌
  - conceptually incoherent to have preferences that are just about reciprocity ... ⇌
  - people **must** have some notion of good and bad outcomes in order to be reciprocal about anything. ⇌
  - Rabin (1993), Dufwenberg-Kirchsteiger embed bad distributional⇌
  - Falk-Fischbacher, Charness-Rabin combine more serious distributional preferences with reciprocity.⇌

↬

Social Preferences

How to test for the role of intentions? $\rightleftharpoons$

- Very simple method:$\rightleftharpoons$
  - With common knowledge to the players, Player A chooses between outcome X and giving Player B the choice from $\{Y, Z\}$.$\rightleftharpoons$
  - Player B's preferences between Y and Z across situations, or when Player A has no choice, reflects B's distributional preferences.$\rightleftharpoons$
  - But the way Player B's preferences between Y and Z depend on changes in X reflect his reciprocal/intentions-based preferences.

$\hookrightarrow$

Social Preferences

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

- Behavior that hurts some or all without helping anybody.⇌

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

⇌

- Note: discussing only B behavior.⇌
  - A behavior in papers.⇌
- Reminder: these data not typical evidence in the literature, where more Pareto damage is typically observed.

↬

Social Preferences

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%*

$\rightleftharpoons$

Increase is statistically significant, but not large. Note that seems that B does not punish A for not wanting to be on the short end herself.

$\hookrightarrow$

Social Preferences

When B makes choices where $ for A at stake, but not $ for B, 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?⇌
- Surprisingly little punishment, even when free. ⇌
  - A puzzle.⇌
- **But behindness aversion virtually vanishes when A has been kind.**

↬

Social Preferences

Crude summary of accumulated experimental evidence:$\rightleftharpoons$

- Not much Pareto-damage without reciprocity $\rightleftharpoons$
- Increase in Pareto-damage by B if A is mean/unfair$\rightleftharpoons$
- Stronger indication of diminishing Pareto-damage if A behaves nicely.

$\looparrowright$

Social Preferences

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±50,0)*			*37%*		*63%*	⇌
	(550,550)			*11%*		*89%*	

- Comment on 1st vs. 2nd line: ⇌
  - Whoa!⇌
- Lots and lots of data since we didn't believe the results. ⇌
  - It's robust. ⇌
  - Statistically significantly in opposite direction as positive reciprocity.

↬

Social Preferences

Accumulated evidence: $\rightleftharpoons$

- Lots of helpful sacrifice.
- *Not* increased by other's good behavior.$\rightleftharpoons$
- But withdrawn if other misbehaves.$\rightleftharpoons$

Some experiments find positive reciprocity, but very few. $\rightleftharpoons$

- My bet: meta-analysis would show people that positive reciprocity in sense of behaving <u>better</u> towards somebody if she has been good is approximately **zero** in the lab. $\rightleftharpoons$
- The "concern withdrawal", sort of in between positive reciprocity and negative reciprocity as conventionally conceived, seems robust.

$\curvearrowright$

Social Preferences

- Explanations, by Bozos like Rabin (1993), that cooperation in laboratory PD is from positive reciprocity, is wrong.$\rightleftharpoons$
- Similarly: massive experiments on "trust" seem really to be about ρ — little indication that trust is rewarded independent of desire to share.$\rightleftharpoons$
- I believe in positive reciprocity$\rightleftharpoons$
  - and have thoughts why missing from lab.$\rightleftharpoons$
  - but it is not in evidence whatsoever in the lab.

$\looparrowright$

Social Preferences

- Approximately:⇌
  - In UG: All negative reciprocity, no behindness aversion.⇌
  - In PD: All equity, no positive reciprocity.⇌
- Over-strong statements? ⇌
  - Stark statements to counter-balance the apparent triumph of priors and theory and pet models over easy-to-see and overwhelming empirical evidence.⇌
- The point is **not** a horse race, ⇌
  - not a claim about which motives bigger.⇌
- Estimates in lab on behindness aversion and positive reciprocity.⇌
- Both are (slightly) backwards.

↬

Social Preferences

Back to belief-based utility?⇌

- Reciprocity models (and cousins/descendants a la Levine "Spite") are clearly belief-based.⇌
- But so are the new models based on image.⇌
- Clearly right direction⇌
  - Even if some of them formulated in complacent terms.⇌
  - And must be "seeded" by distributional assumptions.⇌
  - Because all models must.

↬

Social Preferences

Great early experiment by Dana, Weber, and Kuang (2007):$\rightleftharpoons$

- (other, self) of $A \equiv (\$5, \$5)$ vs. $B \equiv (\$1, \$6)$.$\rightleftharpoons$
  - 26% choose $B \equiv (\$1, \$6)$.$\rightleftharpoons$
- (other, self) of $C \equiv (\$1, \$5)$ vs. $D \equiv (\$5, \$6)$. $\rightleftharpoons$
  - Presumptively, 100% choose D. $\rightleftharpoons$
  - And, in hypothetical treatment, 100% in fact chose it.$\rightleftharpoons$
- Then: people told to choose either \$5 or \$6 for self.$\rightleftharpoons$
  - But 50% chance $(\$\_\_, \$5)$ vs. $(\$\_\_, \$6)$ is A *vs.* B.$\rightleftharpoons$
  - 50% chance it is C *vs.* B.

↬

Social Preferences

- 37% choose to Not reveal, take \$6.$\rightleftharpoons$
- 7% choose to Not reveal, take \$5.$\rightleftharpoons$
- 56% choose to Reveal.$\rightleftharpoons$
- If revealed and saw (\$5,\$5) vs. (\$1,\$6),$\rightleftharpoons$
  - 25% choose (\$1, \$6)$\rightleftharpoons$
- If revealed and saw (\$1,\$5) vs. (\$5,\$6),$\rightleftharpoons$
  - 90% choose (\$5, \$6). (So 10% are either silly or nasty)$\rightleftharpoons$
- And so if choice turned out to be (\$5,\$5) vs. (\$1,\$6), 63% (rather than 26%) ended up choosing (\$1,\$6), either by choosing it after seeing it, or by choosing not to see it.

$\looparrowright$

Social Preferences

Interpretation?⇌

- "Moral Wiggle Room". Authors (incessantly) point out cannot explain by distributional preferences. Why not reveal payoffs?⇌
- One interpretation: self-image preservation.⇌
- Grossman (2008) and Lazear, Malmendier, and Weber (2007), Tadelis, etc.: variants of people seeming to have some variant of belief-based social preferences.
  - Shame? ⇌
  - Social signaling? ⇌
  - Self signaling? ⇌
- People care how obvious it is to themselves and to others that they are being selfish or virtuous.

↬

Social Preferences

Final point .... $\rightleftharpoons$

- "Wiggle room $\longrightarrow$ non-selfish behavior may be less important than it appears" is not an awful intuition.$\rightleftharpoons$
- But it's not as compelling as may seem.$\rightleftharpoons$
- A world where people convince themselves they are right to justify selfish behavior may be *more* different from simple selfishness than is simple simple social preferences.$\rightleftharpoons$
- Bargaining selfish people vs. bargaining altruists vs. bargaining self-righteous people?

$\looparrowright$

Social Preferences

Modeling Reciprocity? Hard.$\rightleftharpoons$

- I mean hard in both the sense of technically difficult and in terms of really finding robust models. Other- or self-signaling use signaling models (and moral wiggle room and related experiments seem to require weirdness; Rabin (1995) "moral rules" vs. "moral preferences" with weird assumption about treatment of information. Reciprocity uses "psychological games" by Geanakoplos, Pearce, and Stacchetti (1989), or related approaches, such as Levine's signaling approach or steady-state equilibrium a la Charness & Rabin.$\rightleftharpoons$
- Technical issues in modeling reciprocity: Both as psychological reality and as modeling strategy, not just outcomes, nor even cleverly defined to include game, but beliefs.$\rightleftharpoons$
- How might we capture the role of volition and intentions in models of social preferences?

$\hookrightarrow$

Social Preferences

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

- Nor, to slightly generalize, can capture by writing payoffs solely function of outcomes.$\rightleftharpoons$
- Point is not that you can't do in terms of own payoff. Duh. The point is that it is not a function of everybody's payoff ... but different things. In any framework, psychological games of type-signaling games, need beliefs in there.$\rightleftharpoons$
- Think about the following game. Fully hypothetical; no data. Going to 3-person for ease of illustration.

↬

Social Preferences

- Want you to think about the following game (using intuition, like I am, since nobody has run the game to my knowledge). Going to 3-person for ease of illustration.⇌

	1 goes L		1 goes R	
	2 chooses		3 chooses	
	L	R	L	R
1's Payoff	10	12	11	5
2's Payoff	10	12	5	5
3's Payoff	10	0	5	5

⇌

What do you think Player 3 will do if Player 1 goes R? What are the issues?⇌

- If Player 3 thinks Player 2 would have gone L?⇌
- If Player 3 thinks Player 2 would have gone R?⇌
- What is not right about those questions?

Introduction to Limited Rationality

Are people fully rational in maximizing their utility?⇌

- No.⇌

Huge variance in tolerance for doing models of it.⇌

- Bounded rationality long been a topic of economic research.⇌
- And (especially recently) great interest to economic theorists.⇌

Two claims:⇌

1. It is obvious people are not 100% rational, and that they are not 100% in commonly occuring high-stakes settings.⇌
2. And obvious that departures from 100% rationality matter.

↩

Introduction to Limited Rationality

Proof of Claim 1:$\rightleftharpoons$

- Not getting A's in undergrad economics courses, or 800 on math and analytic SATs, GREs, etc., are departures from rationality with huge costs for those who make them.$\rightleftharpoons$
- The "real world" \$ cost to a smarter-than-average student at an elite university commiting the sunk-cost fallacy on an exam may be higher than most commiting it in the field, and yet many do so.$\rightleftharpoons$
- Benefit of solving unsolved problems in auction theory higher (instant lucrative tenure at top schools) than to MBAs in charge of bidding.$\rightleftharpoons$

Proof of Claim 2:$\rightleftharpoons$

- Billions of \$ in direct educational costs and maybe trillions of \$ on opportunity costs to teach math.

$\hookleftarrow$

Introduction

- Approaching limited rationality in economics:

2. Show what limits explain, not what explains limits.⇌

- Often nothing explains it.⇌
- More importantly,⇌
  - It's not the traditional focus of economics.⇌
  - I think it *shouldn't* be a major focus of economics.⇌
  - Uncontroversially, should not be *unique* focus of economics.⇌

3. Understand human irrationality is not just about complexity.⇌

- Economists drawn to complexity explanations because:⇌
  - It's realism: intuitive appeal and reasonableness.⇌
- Attraction to computer science and math.⇌
- "Cognitive miser" appealing construct.

↬

Introduction

Human irrationality is not just about complexity.⇌

- Economists drawn to complexity explanations because:⇌
  - It's manifest realism: intuitive appeal and reasonableness.⇌
  - Attraction to computer science and math.⇌
  - "Cognitive miser" appealing construct.

Lesson from psychology that many important errors:⇌

- Not from fact that getting the right answer is so hard⇌
- But because the wrong answer is so enticing⇌

"Bound Errors" vs "Astray Errors"⇌

- Realizing you'll have same self-control problem tomorrow as today isn't cognitively harder than thinking it'll be different.⇌
  - Convoluted, time-consuming stories for why we'll be better ...⇌
  - 1,000 reasons didn't start your diet or dissertation ...⇌ but tomorrow.⇌
- Planning future as if current craving state will last forever⇌
  - More complicated than using average!

↬

Introduction

A lot of cognitive energy recognizing patterns that are not there.⇌

- "covariation bias"⇌ (minimal dectable correlation?)⇌
- gambler's fallacy, hot-hand fallacy, etc.⇌
  - What you should think the next flip of a coin is:⇌ 50-50.⇌ Done.⇌
  - Instead: people think next flip depends on recent flips.⇌

Investors lose **lots** of money because don't use the simplest investment strategy of putting money in index fund.⇌

- See predictable patterns in stock markets that are not there.⇌
- Don't ask the question: "Why is this person trading with me?"⇌
- If *either* thought a lot less, or asked that one simple question more, would save lots of money.

↬

Some Themes/Perspectives on Limited Rationality$\rightleftharpoons$

Right that much bounded rationality seems "constrained optimal".$\rightleftharpoons$

- Given that our brains are not perfect, we make do with approximate heuristics; conserve energy, don't pay infinite attention, etc.$\rightleftharpoons$
- Identifying the source of "bounds" in bounded rationality does not magically make it unboundedly rational.$\rightleftharpoons$
- We won't emphasize the "theory-office neuro" that seeks to find the "foundations" of the human error as a constrained-maximization solution to structures that brains (are imagined to) have.

$\hookleftarrow$

Introduction to Limited Rationality

Labeling does not matter.$\rightleftharpoons$

- Not paying attention to details might legitly be called rational.$\rightleftharpoons$

But never use labeling to obfuscate:$\rightleftharpoons$

- If consumers' unenthusiasm for detailed calculations to choose optimal phone contract is called rational, do not then study markets or policy as if consumers are "no-mistake rationals".

And biggest theme of all:
$\hookleftarrow$

Introduction to Limited Rationality

"Bound errors" and "Astray errors" are often complements.⇌

- In a complicated and distracted world, don't/can't solve.⇌
  - But Solow's Law: can't replace "rational" with "not rational".⇌
- So when you're guessing, what intuitions lead you astray?⇌
  - If can't solve for optimal auction bid, crudely adjust for WC?⇌

Other themes in models of limited rationality:⇌

- **The "Slander Paradox" of modeling limited rationality**⇌
  - We write down simple models so that *we* can solve.⇌
  - But then make agents look like idiots.

↩

Introduction to Limited Rationality

Whether people get right answer vs. whether they "know" right answer?⇌

- Two cautions:⇌
- Often smarter in fact than in principle/abstract:⇌
  - Per Friedman's famous illustration, an expert may be brilliant at pool without knowing the laws of physics.⇌
  - Tacit knowledge: economic actors often closer in practiced settings to idealized model of superhuman rationality than appears in direct tests.⇌
- But also: often smarter in principle/abstract than in fact.⇌
  - **Errors of implementation**.⇌
  - Thaler playing pool.⇌
  - You are masters of probability theory and would not make errors when the probability theory is made salient ... but commit those errors in life.

↩P

Introduction to Limited Rationality

Useful to conceptualize Limited Rationality into two categories:

1. "Quasi-Maximization" $\rightleftharpoons$
   1. Present Bias $\rightleftharpoons$
   2. Preference Misprediction $\rightleftharpoons$
   3. Narrow Bracketing and Narrow Focusing $\rightleftharpoons$
2. "Quasi-Bayesian". $\rightleftharpoons$
   1. Errors in probabilistic judgments about things $\rightleftharpoons$ ... base-rate neglect, gambler's fallacy, $\rightleftharpoons$
   2. motivated cognition: preferences and emotions distorting probabilistic judgments $\rightleftharpoons$ ... cognitive dissonance, self-serving biases $\rightleftharpoons$ (too little at this particular camp) $\rightleftharpoons$
   3. Errors in statistical reasoning about volitional agents $\rightleftharpoons$ ...level-k reasoning, cursedness and inferential naivety, hindsight bias

$\hookleftarrow$

Introduction to Limited Rationality

"Quasi-Maximization": Model person as engaging in traditional constrained maximization at each moment in time.⇌

- But specify exact mistake the person is making in which function she is maximizing, or in what choice set she is choosing from.⇌

Does not correspond to maximizing *true* preferences because ...⇌

- Present bias: moment by moment, you maximize full intertemporal utility, but at each moment tend to overweight current utility.⇌
  - And may mispredict the propensity to do so in the future.⇌
- Utility misprediction: because of current tastes or current focus, you (actively or passively) mispredict utility of future situations⇌
  - misremembering past episodes.⇌
- "Decision neglect" and "narrow bracketing": maximizing true utility among each choice set you focus on, but don't focus globally⇌
  - don't think about some choices, and don't integrate others.

Introduction to Limited Rationality

Lesson and theme for economics of quasi-maximization perspective:⇌

- Have we ever chosen our "life course"?⇌
- Our piecemeal maximization may lead to life course we never *chose*.⇌
  - A smoker 'decided' thousands of times to smoke ...⇌ but did she ever decide to become a smoker?⇌
  - An overweight person 'decided' thousands of times to eat more calories than he burned⇌ ... but ever decide to become an overweight person?⇌
  - A person in $32,000 credit-card debt made all the choices leading to that debt⇌ ... but did she ever decide to be $32,000 in debt?⇌
- All 3 Quasi-Max errors contribute.

↩

Introduction to Limited Rationality

Person maximizes a particular "goal" given his choice set:$\rightleftharpoons$

$$\text{Max}_{x \in X} \; V(x).\rightleftharpoons$$

But V not actual utility function should be maximizing.

$\hookleftarrow$

Mispredicting Tastes

When making choices, we must (or should) take into account any impact current choices have on future well-being.$\rightleftharpoons$

- Apart from question of **how much she cares** about future utility...$\rightleftharpoons$ question of **what she believes** about future utility.

$\hookleftarrow$

Mispredicting Tastes

Note some issues missing because: ignoring belief-based utility⇌

- Mistaken beliefs about future tastes can matter both now and in the future *directly*–even if it does not change behavior. ⇌
- If going to a restaurant a week from now.⇌
  - enjoyment of meal depends in part on whether disappointed. ⇌
  - Hence, beliefs right before the meal matter - whether right or wrong.⇌
  - And if anticipation matters, can't depend on correctness.

↩℘

Mispredicting Tastes

Basic approach:$\rightleftharpoons$

- People have perceptions what will make them happy, take actions they assess as maximizing this perceived happiness.$\rightleftharpoons$
- But perceptions may be wrong.$\rightleftharpoons$

Formally: true utility is given by: $U^t = \sum_{\tau=t}^{T} u(\mathbf{c}_\tau, \mathbf{s}_\tau)$,$\rightleftharpoons$

- where $u(\mathbf{c}_\tau, \mathbf{s}_\tau)$ is her instantaneous utility in period τ,$\rightleftharpoons$
- T is her (possibly infinite) time horizon,$\rightleftharpoons$
- $\mathbf{c}_\tau$ is period-τ consumption vector $\rightleftharpoons$
  - includes all period-τ behavior relevant for contemporaneous utility.$\rightleftharpoons$
- Vector $\mathbf{s}_\tau$ is person's "state" in period τ, which incorporates all factors that affect instantaneous utility besides current consumption.$\rightleftharpoons$

$\hookleftarrow$

Mispredicting Tastes

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

- θ_t is all things that might affect predictions besides the current
- θ_t will always contain $\mathbf{s}_t$, which (we shall shortly see) may be a major influence on prediction about the future.

Conventional assumption:

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

But we allow:

- Systematic factors in θ_t that influence prediction of future utility?

↩

Mispredicting Tastes

These predictions ought be probabilistic:$\rightleftharpoons$

- person may have non-firm beliefs about her future utility.$\rightleftharpoons$
- We'll ignore.$\rightleftharpoons$
- People maximize $\widetilde{U}^t$ rather than U.$\rightleftharpoons$
- Behavioral implications$\rightleftharpoons$
- Because $\tilde{U}^t$ can vary by current state.$\rightleftharpoons$
  - people may exhibit *dynamic inconsistency*$\rightleftharpoons$
    - changing plans even in absence of new information.

$\hookleftarrow$

Mispredicting Tastes

Models of such bias?⇌

- Only formal theory I know of is "projection bias".⇌
  - Loewenstein, O'Donoghue, and Rabin (2003)⇌
  - (But see recent formal model by Bushong and Gagnon-Bartsch on attribution bias)
- In strongest and simplest form:⇌
  - $\tilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau | \mathbf{s}_t) = u(\mathbf{c}_\tau, \mathbf{s}_t)$.⇌
- People underappreciate (even predictable) changes in tastes.⇌
  - And hence falsely project their current tastes onto the future.⇌
- LOR claim general error embeds lots of intuitions:⇌
  - Kahneman's "transition heuristic"⇌
  - Gilbert's "Immune Neglect"⇌
  - (An interpretation of) Herrnstein and Prelec's "melioration"

↩

Mispredicting Tastes

Language and Psychology:$\rightleftharpoons$

- Misprediction vs. underappreciation:$\rightleftharpoons$
  - Caveat to the language and formalization using here:$\rightleftharpoons$
  - Is it really always a sort of "active misprediction" in the sense that people have articulated beliefs, but they get them wrong?

$\hookleftarrow$

Mispredicting Tastes

Two ways tastes change over time:⇌

- Temporary Fluctuations⇌
  - cues,⇌ satiation and deprivation,⇌ random moods,⇌ etc.⇌
- Longer-term changes⇌
  - adaptation—paraplegia, standard of living, etc., virtually any life event, good or bad, that we tend to adapt to.⇌ addiction,⇌ etc.⇌

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

- The fact that tastes change, while not regularly part of core economics, is utterly consistent with rational choice, and has been much studied in rational-choice models recently.

↩

Mispredicting Tastes

10 domains of evidence so far (in increasing order of importance)$\rightleftharpoons$

1. Clothing$\rightleftharpoons$
2. Convertibles$\rightleftharpoons$
3. Swimming Pools$\rightleftharpoons$
4. Winning Lotteries$\rightleftharpoons$
5. Paraplegia$\rightleftharpoons$
6. Food$\rightleftharpoons$
7. Sex$\rightleftharpoons$
8. Drugs$\rightleftharpoons$
9. Academic tenure$\rightleftharpoons$
10. Mugs

$\hookleftarrow$

Mispredicting Tastes

Can identify misprediction of **marginal** utility using...$\rightleftharpoons$ price theory.$\rightleftharpoons$

- How habit forming cigarettes are.$\rightleftharpoons$
- If WTP for something in period t is systematically different in period $t' \leq t$ vs. period $t'' \leq t$, evidence misprediction.$\rightleftharpoons$
- (If $t', t'' < t$, unconfounded by present bias)$\rightleftharpoons$

But identifying misprediction of **levels** of utility is much harder.$\rightleftharpoons$

- Identifying mispredictions of how bad we'll feel smoking very hard$\rightleftharpoons$

A major form of evidence for preference misprediction:$\rightleftharpoons$

- value for future consumption depends on *current* situation in ways $\rightleftharpoons$
  - we think "can't" affect actual future value.$\rightleftharpoons$
  - and can't be information$\rightleftharpoons$
- This turns out to be some of the best evidence.

Mispredicting Tastes

Underappreciating Fluctations in Taste

- Our utility for many important and familiar activities in vary greatly.⇌
- Nothing in economics says taste for food, sex, sleep, drugs, mugs (big five!) won't fluctuate greatly based on satiation and circumstances.⇌

We look to these activities for misprediction.⇌

1. Because of their importance, and⇌
2. Because universal experience with fluctuations, mispredictions highly suggestive of important bias, not just limited information.⇌

Despite *vast experience*, we underestimate these fluctuations.

↩

Mispredicting Tastes

Food$\rightleftharpoons$

- Studies support folk wisdom: buy more when shopping on empty stomach. $\rightleftharpoons$
- E.g., Nisbett and Kanouse (1968) Gilbert, Gill and Wilson (1998).$\rightleftharpoons$
- Classic: randomly give people muffins as they enter grocery store ... see if this influences their shopping.$\rightleftharpoons$
- When currently hungry act as if future taste for food will reflect such hunger, and when sated act as if sated in future.$\rightleftharpoons$
- Order too much when sitting down to meal, not realizing you'll become sated (for those of you who ever become sated...).$\rightleftharpoons$
- Or planning diet right after stuffing your face/not dieting.$\rightleftharpoons$
  - Not hungry now, so think can go the rest of your life without eating!

$\hookleftarrow$

Mispredicting Tastes

Evidence? Two beautiful experiments ...⇌

- Food⇌
- Addictive drugs⇌

Read and van Leeuwen (1998):⇌

- Office workers were asked to choose between "healthy snacks" and "unhealthy snacks" that they would receive in one week.⇌
  - Either when they should expect to be hungry (late in afternoon) or⇌
  - satiated (immediately after lunch).⇌
- Healthy: apples and bananas.⇌
- Unhealthy: crisps, borrelnoten, Mars Bars, and Snickers Bars.⇌
- Identifying hypothsis:⇌
  - "Unhealthy" more appealing when hungry.

↩

Mispredicting Tastes

% of Subjects Choosing "Unhealthy" Snack

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

- What does comparing columns tell us?
  - That the identifying hypothesis was right.
- What does comparing the rows tell us?
  - Projection bias.

Mispredicting Tastes

The food misprediction may or may not be important.⇌

- But should be kept in mind in more difficult cases that might more plausibly reflect limited information:⇌
- We experience fluctuations in hunger tens of thousands of times.⇌
  - Same with sleepiness, ⇌
  - sexual arousal,⇌
  - drug craving,⇌
  - mug craving⇌

Reminder of psychological/conceptual/linguistic problem:⇌

- "Misprediction" really the right term?⇌
- Nobody lacks cognitive access to effects of hunger state.⇌
- "Under-attendance-to" might be the right term.

↩

Mispredicting Tastes

Beautiful simple experiment in more dramatic context⇌

- Giordano, Bickel, Loewenstein, Jacobs, Marsch, and Badger (2002), and another permutation:⇌

Elicited (real stakes) WTP for *second* dose of heroin substitute BUP from 13 long-time heroin addicts regularly receiving single dose.⇌

- Reduces craving for heroin, aiding withdrawal⇌
- Asked: would you rather have⇌
  - 2nd dose or \$10?⇌
  - 2nd dose or \$20?⇌
  - ...⇌
  - 2nd dose or \$100?⇌
- Subjects told (truthfully) choice randomly implemented.⇌
- **Only 13 subjects!**

↩

Mispredicting Tastes

Could not legally or ethically withhold the usual first dose, which in fact was usually the only dose given.$\rightleftharpoons$

- But this 2nd dose is still attractive to addicts.$\rightleftharpoons$

Half asked when "more deprived" about 2nd dose, half asked when "less deprived" about 2nd dose.$\rightleftharpoons$

- More deprived: 2 hours before scheduled dose.$\rightleftharpoons$
- Less deprived: right after scheduled dose.$\rightleftharpoons$

Half asked WTP for 2nd dose today, half about 2nd dose on next visit.$\rightleftharpoons$

- Observe: always being asked about a *second* dose.

$\hookleftarrow$

Mispredicting Tastes

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 *exact same circumstances by experienced* addicts.⇌

- Note the present bias.⇌
- But also note the projection bias.⇌

What is the "true" value?⇌

- Almost surely the \$35.⇌
- Why?

↬

Mispredicting Tastes

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

By BE principles ... probably the \$35 is closest to true well-being.. *Not* because drugs are bad and this is lowest ...but because

1. present bias says 2nd column better than first, and
2. projection bias says bottom row better than top.

↬

Mispredicting Tastes

Not a reckless act of proto-fascism to guess so.⇌

- Not psychologists, behavioral economists, government officials, doctors, or preachers declaring that people are too keen on drugs.⇌

4 different answers to what theory had said is same question⇌

- Incoherent to "use addict's WTP" when he has 4 different WTP's.⇌

Two rules of thumb:

- Don't follow normative advice from those descriptively wrong.
- If claim choice is error, must be **some** way to elicit what you believe is right choice.
- (Neither universal)

↬

Mispredicting Tastes

Hypothetical experiment:

		When they would get the dose	
		Today	Next visit
Current Craving	Deprived	$250	$200
	Satiated	$150	$100

- Right answer probably $200, not $150 or $100. Nor "$0".
- Maybe moral or externality reasons, etc.
- But not the mistakes we have explored.

Cautionary tale:

- Present bias makes us advocate prospective choice.
- But prospective choice bad if in different state
  - Very often **is** in different (cold) state.
- We can be amazingly un-empathetic & cruel to future craving self!
  - Being so cruel is a mistake.

Mispredicting Tastes

Field evidence?$\rightleftharpoons$

- Levy (2008) on tobacco$\rightleftharpoons$
- Acland and Levy (2010): gym-goers in an incentivized experiment do not appreciate the positive addiction of exercise regimes.$\rightleftharpoons$
- Conlin, O'Donoghue, and Vogelsang (2007): use returns of cold-weather clothing to estimate projection of sensations of being cold onto future.$\rightleftharpoons$
- Busse, Pope, Pope, and Jorge Silva-Risso (2012) on cars (and car roofs) and houses (and swimming pools).$\rightleftharpoons$
- Augenblick and Rabin (2015) on willingness to do unpleasant task in future$\rightleftharpoons$
  - Decide before or after you've just done it?

Mispredicting Tastes

Augenblick and Rabin (2018): $\rightleftharpoons$

- Estimate β (present bias), β_h (sophistication over present bias), and α (projection bias).$\rightleftharpoons$
- Design: $\rightleftharpoons$
  - People choose over effortful tasks at different wages and different dates.$\rightleftharpoons$
  - People also predict (in incentivized way) what they would choose at future dates and wages.$\rightleftharpoons$
- Difference between tasks completed today and in future identifies β.$\rightleftharpoons$
- Difference between prediction and actual effort identifies β_h (under caveat of consistency preferences).$\rightleftharpoons$
- **Difference between tasks chosen before and after completing mandatory work identifies α.**

↬

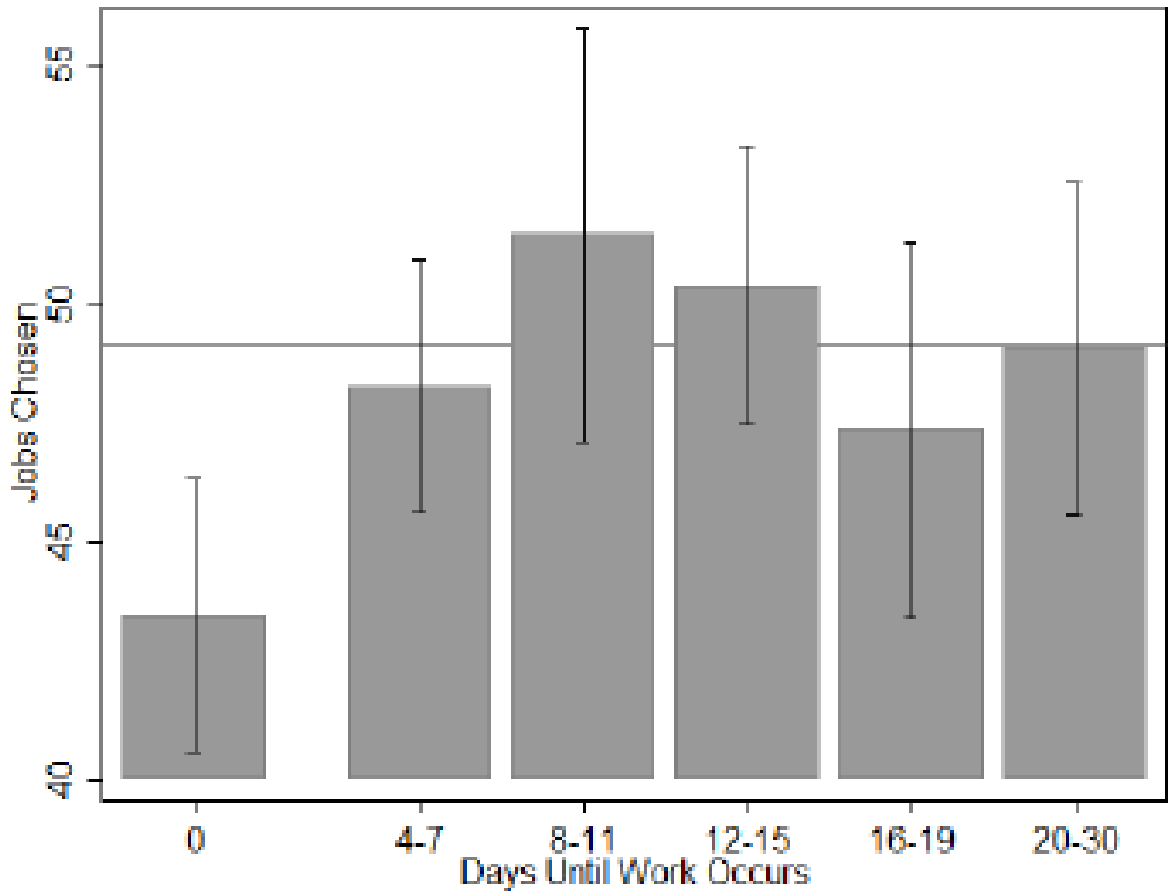
Mmmm... Salty

Skipping over components, $\rightleftharpoons$

- Task is transcription of blurry greek letters, while getting pestered by noise requiring button press.$\rightleftharpoons$
- Each Day: **Complete mandatory work: 10 mandatory tasks.**$\rightleftharpoons$
- Stating preferences over immediate (today's) work.$\rightleftharpoons$
- Stating preferences over 2 future dates (except on last day).$\rightleftharpoons$
- Stating predictions over possible future work (except on last day).$\rightleftharpoons$
- Observing supplemental work decision that is implemented today.$\rightleftharpoons$
- Completing supplemental work decision

$\looparrowright$

Mispredicting Tastes



↬

Mispredicting Tastes

Overview of results:

- $\beta \in (0.812, 0.835)$
- $\beta_h \in (0.999, 1.014)$
- $\delta \in (1.003, 1.005)_{\rightleftharpoons}$

Update ... Augenblick follow-up on 1-4 day delay ...

$\looparrowright$

Mispredicting Tastes

Relevant Part Here: $\rightleftharpoons$

- After completing mandatory tasks, tasks completed today decrease by $2-3$.$\rightleftharpoons$
- Taking censorship into account (40% of participants choose 0 or 100 tasks), this increases to $4-6$ depending on specification.$\rightleftharpoons$
- Held for immediate work, future work, and prediction of future work$\rightleftharpoons$
  - All as predicted by projection bias$\rightleftharpoons$
  - Note also: not a decrease over the weeks.

$\looparrowright$

Mispredicting Tastes

Busse, Pope, Pope, and Russo: test for projection bias (PB) in housing and car market⇌

- weather impacts current perception of value of⇌
  - convertibles, 4-wheel drives, black cars⇌
  - swimming pools and central air in homes

↬

Mispredicting Tastes

Data is ~20% of new car dealerships in US from January 1st 2001 to December 31st 2008⇌

- Includes date and location of purchase, some car characteristics, and price paid⇌
  - Locations are Nielsen Designated Market Areas (DMAs)⇌

Results for convertibles:⇌

- Huge seasonal fluctuations (as expected)⇌
- 10° Fahrenheit higher (than average in that DMA in that week of the year) leads to 0.07% points more⇌
  - Increase of 2.7% on base rate of 2.6%.⇌
- Going from completely clear to completely covered sky leads to a 0.126% points decrease⇌
  - Equivalent to roughly 18 degree Fahrenheit change⇌
- Same regression for 4-wheel drives yields similar results⇌
- Careful about alternative stories that might be driving the results

↬

Mispredicting Tastes

They also follow Conlin et al on clothing: $\rightleftharpoons$

- Do people return "mistakenly bought" convertibles more often?$\rightleftharpoons$
  - Maybe a little.$\rightleftharpoons$
- Do people return "mistakenly bought" 4-wheel drives more often?$\rightleftharpoons$
  - Tentatively yes.$\rightleftharpoons$

Housing market and swimming pools from NBER working paper, NOT from QJE paper.$\rightleftharpoons$

- Data: 4 million single-family residential properties across the US that sold at least twice in 1998-2008.$\rightleftharpoons$
- People pay more for pools and AC when weather hot.$\rightleftharpoons$
- And non-random weather variation more of puzzle:$\rightleftharpoons$
  - People get pools when *buy* in very hot month of August, though get house in October.

Mispredicting Tastes

Longer-Term Changes$\rightleftharpoons$

- We adapt to changes.$\rightleftharpoons$
- We form habits.$\rightleftharpoons$

Do we understand such changes?$\rightleftharpoons$

- We underestimate adaptation.$\rightleftharpoons$
- We underestimate habit formation.

$\hookleftarrow$

Mispredicting Tastes

Suppose each day of your life choosing whether to sniff widgets.

- $a_t = y$ if sniff, and $a_t = n$ if don't sniff.

Sniffing has the following utility function:

$$u_t(a_t = y | a_{t-1} = n) = 5$$
$$u_t(a_t = n | a_{t-1} = n) = 0$$
$$u_t(a_t = y | a_{t-1} = y) = -2$$
$$u_t(a_t = n | a_{t-1} = y) = -12$$

Some features of widget sniffing?

- You always enjoy sniffing
- Sniffing lowers future utility (from both sniffing and not sniffing)
- But it raises the *marginal* utility of sniffing in the future.
- Optimal behavior? Projection-biased behavior?
- Aside: Sophisticated present bias? Naive present bias?

Mispredicting Tastes

Evidence on mispredicting longer-term changes? $\rightleftharpoons$

- Much harder ...$\rightleftharpoons$

Hint of evidence on habit formation (!), and of direct interest:$\rightleftharpoons$

- Projection bias and Loss Aversion.

$\looparrowright$

Mispredicting Tastes

Classic study by Brickman, Coates, and Janoff-Bulman (1978):⇌

- People underappreciate adaptation—winning lotteries, losing limbs.⇌
- Interviewed people who had won lottery jackpot prizes within the last year (average winnings of $479,545) and a control group.⇌
- Virtually no difference in reported happiness.⇌
- No data on non-winners' predictions of how they would feel if they won.⇌
- But the notion that lottery winners are no happier than non-winners surely runs counter to the predictions of most people.⇌
  - Including, presumably, those playing the lottery.⇌
- Would people play if predicted full adaptation?⇌
- Some suggestive stronger evidence:

↬

Mispredicting Tastes

Cohen (1999, unpublished), cited by Kahneman (2000), asked a bunch of people to predict: "What % of time do you think somebody will be in Good ... Bad ... Neutral Mood" Y months after event X?⇌ She also asked respondents whether they knew somebody with that experience.⇌

Event:	Paraplegic		Lottery Winner	
Know somebody?	No	Yes	No	Yes
%G-%B One month later	-41	-50	58	64
%G-%B One year later	-37	-19	50	25

⇌

- How prediction immediate differ?⇌
- Prediction a year later?⇌

Hint of evidence on adaptation, and of direct interest:⇌

- Projection bias and Loss Aversion.

↩

Mispredicting Tastes

- **Behavioral economists love mugs ...**⇌
  - It's like heroin or borrelnoten to us!⇌
- Do people predict loss aversion?⇌
- Loewenstein and Adler (1995) randomly assigned subjects.⇌
- "Prediction" treatment 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.⇌
- Shown form to be used to elicit their selling price, asked to complete it as they expected they would once they received the mug.⇌
- Then given mug, and form eliciting selling prices.⇌
- Other half of subjects given mugs without first making predictions, then completed the form eliciting selling prices.

↩

Mispredicting Tastes

Condition	Prediction of Valuation	Actual Valuation
Prediction	\$3.45	\$4.89
No Prediction	———	\$5.56

Interpretation: mini-look at habit formation.

- People underappreciate how ownership will affect their valuation.

Question: Why did they do the "no-prediction" condition?

- Answer: worries about dissonance/consistency

Question: What do the results indicate?

- Consistency indeed had some implications.

Mispredicting Tastes

An aside: is loss aversion itself a mistake?$\rightleftharpoons$

- Loss aversion clearly a real hedonic experience people have.$\rightleftharpoons$
- Incoherent virtually to interpret behavior otherwise.$\rightleftharpoons$

But two types of mistakes mean people over-attentive to gains and losses, so significantly a mistake.$\rightleftharpoons$

- Projection bias suggests that people over-attentive to losses and to gains, since don't realize that the sensation will wear off.$\rightleftharpoons$
- Next up: narrow bracketing makes people over-attentive to losses and gains, since don't think about how it will cancel out.

Mispredicting Tastes

Mugs Model worked out

- Old-fashioned view of loss aversion as involving status quo, not expectations. $\rightleftharpoons$
- Similar lessons surely hold if one models projection bias over belief-based preferences, but we have no formal models of that. $\rightleftharpoons$
- In period t a person can either "consume" a mug ($c_t = 1$) or not ($c_t = 0$). $\rightleftharpoons$
- Person can either feel endowed with mug ($s_t = 1$) or feel unendowed with mug ($s_t = 0$). $\rightleftharpoons$
- $s_t = c_{t-1}$

$\hookrightarrow$

Mispredicting Tastes

In experiment:$\rightleftharpoons$

- If the person decides to possess the mug, then $c_1 = c_2 = 1$.$\rightleftharpoons$
- If decides not to possess the mug, then $c_1 = c_2 = 0$.$\rightleftharpoons$

Both sellers and buyers choose between these two consumption flows$\rightleftharpoons$

- But sellers feel endowed in period 1 — they have $s_1 = 1$. $\rightleftharpoons$
- Buyers feel unendowed — they have $s_1 = 0$.$\rightleftharpoons$
- If in period 1 the person chooses to possess the mug, she will feel endowed in period 2 — she'll have $s_2 = 1$.$\rightleftharpoons$
- If in period 1 the person chooses not to possess the mug, she'll have $s_2 = 0$.

$\looparrowright$

Mispredicting Tastes

True Utilities:

u_t(mug now?| mug last period?)		consumption u		gain-loss u
u_t(y|y)	=	1	+	0
u_t(n|y)	=	0	+	-λk
u_t(y|n)	=	1	+	k
u_t(n|n)	=	0	+	0

where $y = 1$ and $n = 0$.

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

Consider the two extreme cases:

$$\alpha = 0 \Longrightarrow \textit{Rational expectations} \Longrightarrow \tilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau | \mathbf{s}_t) = u(\mathbf{c}_\tau, \mathbf{s}_\tau)$$

$$\alpha = 1 \Longrightarrow \textit{100\% projection bias.} \Longrightarrow \tilde{u}(\mathbf{c}_\tau, \mathbf{s}_\tau | \mathbf{s}_t) = u(\mathbf{c}_\tau, \mathbf{s}_t)$$

Mispredicting Tastes

Rational Buying Price:$\rightleftharpoons$

- $B^{RAT} = u(y|n) - u(n|n) + \sum_{t=1}^{\infty} \delta^t [u(y|y) - u(n|n)]$
- $= 1 + k + \frac{\delta}{1-\delta}(1).\rightleftharpoons$

Rational Selling Price:$\rightleftharpoons$

- $S^{RAT} = u(y|y) - u(n|y) + \sum_{t=1}^{\infty} \delta^t [u(y|y) - u(n|n)]$
- $= 1 + \lambda k + \frac{\delta}{1-\delta}(1).\rightleftharpoons$

So: $L^{RAT} \equiv \frac{S^{RAT}}{B^{RAT}} = \frac{1+\lambda k(1-\delta)}{1+k(1-\delta)} > 1.\rightleftharpoons$

- Observe: $\frac{\partial L^{RAT}}{\partial \lambda} > 0$, $\frac{\partial L^{RAT}}{\partial k} > 0$, $\frac{\partial L^{RAT}}{\partial \delta} < 0.\rightleftharpoons$
  - Interpret?$\rightleftharpoons$
- Observe: $Lim_{\delta \longrightarrow 1} L^{RAT} = 1.\rightleftharpoons$
  - Interpret?

Mispredicting Tastes

Projection-Bias Prices (exercise: derive these):$\rightleftharpoons$

$$B^{PROJ} = u(y|n) - u(n|n) + \sum_{t=1}^{\infty} \delta^t [u(y|n) - u(n|n)] =$$
$$1 + k + \frac{\delta}{1-\delta}(1+k).\rightleftharpoons$$

$$S^{PROJ} = u(y|y) - u(n|y) + \sum_{t=1}^{\infty} \delta^t [u(y|y) - u(n|y)] =$$
$$1 + \lambda k + \frac{\delta}{1-\delta}(1+\lambda k).$$

$\looparrowright$

Mispredicting Tastes

Observe: $B^{PROJ} > B^{RAT}$, $S^{PROJ} > S^{RAT}$.$\rightleftharpoons$

Observe: $L^{PROJ} \equiv \frac{S^{PROJ}}{B^{PROJ}} = \frac{1+\lambda k}{1+k}$.$\rightleftharpoons$

Observe: $L^{PROJ} > L^{RAT}$.$\rightleftharpoons$

So an interpretation of the evidence: $\rightleftharpoons$

- Loss aversion rationally derive from real RD-VNM prefs, but by narrow bracketing and projection bias are irrationally magnified.$\rightleftharpoons$

Can projection bias explain the Loewenstein and Adler stuff? $\rightleftharpoons$

- We know $S^{PROJ} = \sum_{t=0}^{\infty} \delta^t [u(y|y) - u(n|y)] = \frac{1}{1-\delta}(1+\lambda k)$.$\rightleftharpoons$
- But what will be unendowed Person's Prediction, $E[S^{PROJ}|n]$? $\rightleftharpoons$
- Claim: $E[S^{PROJ}|n] \equiv B^{PROJ}$ $(< S^{PROJ})$.

$\hookleftarrow$

Mispredicting Tastes

Melioration$\rightleftharpoons$

- My spin (not theirs, and not capturing everything): Herrnstein and Prelec's notion of "melioration" is projection bias.$\rightleftharpoons$
- Every day Theobald goes to eat in one of the only two restaurants in Scagsville: a (Austrian cuisine) or b (Belarusian cuisine).$\rightleftharpoons$
- Theobald is completely patient ($\delta \longrightarrow 1$) and will live forever.$\rightleftharpoons$
- His utility each day t is determined entirely by where he eats day t, $f_t \in \{a, b\}$, and where he ate yesterday, $f_{t-1} \in \{a, b\}$.

$\leftrightarrow\!\!\!\!\!\rho$

Mispredicting Tastes

His instantaneous utility on date t is given as follows:

$$u_t(f_t = a | f_{t-1} = a) = 6$$
$$u_t(f_t = b | f_{t-1} = a) = 5$$

$$u_t(f_t = a | f_{t-1} = b) = 9$$
$$u_t(f_t = b | f_{t-1} = b) = 4$$

Intuition for these preferences?

- Theobald always likes Austrian food better than Belarusian
- He likes each cuisine better when he's not eaten it recently.

Question: How Theobald maximize total utility from meals over life?

- He should alternative between Austrian and Belarusian

Why?

- Because that gives him average of $(5+9)/2 = 7$, rather than average of 6 if he always eats Austrian.

Mispredicting Tastes

$u_t(f_t = a | f_{t-1} = a) = 6$
$u_t(f_t = b | f_{t-1} = a) = 5$

$u_t(f_t = a | f_{t-1} = b) = 9$
$u_t(f_t = b | f_{t-1} = b) = 4$

Will Theobald alternate thusly?

- We just said yes if fully rational.
- No if myopic in sense of taste for immediate gratification.
  - Unlikely
  - Exercise: for what β, δ combinations will Theobald always eat Austrian?
- But "no" for more likely reason:
  - Don't see through logic of how current behavior affects future utility.
  - Projection bias makes you ignore the "internalities".
  - Different than not *caring* about future utility.
- Stress test: is projection bias calibrationally plausible for internality neglect?

Mispredicting Tastes

Intrapersonal vs. Interpersonal Preference (Mis)prediction⇌

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

- Sellers of mugs or houses project their own tastes onto others ⇌
- old people project their (current) tastes on teens.⇌

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

- Tobacco addicts know better how non-addicts will feel when they become addicted than the non-addicts do.⇌

Accuracy may depend on current state of predicter, not their identity.

↬

Mispredicting Tastes

Interpersonal Projection Bias$\rightleftharpoons$

- We could formalize ability and inability to understand others' preferences as "Interpersonal Projection Bias"—biased assessment of others' preferences based on our own *current* preferences.$\rightleftharpoons$
- Importance of interpersonal projection bias?$\rightleftharpoons$
  - A claim (relevant to issues in how economists go about making welfare judgments or refusing to do so): For two people j and k, may often be that: $\tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k | \mathbf{s}_t^j = \mathbf{s}_\tau^k)$ is more accurate than $\tilde{u}(\mathbf{c}_\tau^k, \mathbf{s}_\tau^k | \mathbf{s}_t^k \neq \mathbf{s}_\tau^k)$.$\rightleftharpoons$
- Interpersonal empathy gaps may be profoundly important.$\rightleftharpoons$
  - Much of our lives are understanding motives of others.$\rightleftharpoons$
  - Sometimes it requires imagination about their (different) preferences.$\rightleftharpoons$

$\hookleftarrow$

Mispredicting Tastes

But if we generically have different tastes?⇌

- We may underappreciate how others differ.⇌

Gagnon-Bartsch (2013): ⇌

- social inference if we project own tastes on others.⇌

His favorite example:

- People who dislike gym can't imagine those (freaks) who like the gym.⇌
- Exaggerate health benefits to explain...⇌
- So go to the gym

↩

Mispredicting Tastes

Errors because of such mispredictions?⇌

Many health officials, researchers, and community activists think the sexual-arousal and alcohol-arousal mispredictions are huge issue:⇌

- Very hard to get people to fully appreciate how they are likely to behave later that night when going out.⇌
- Underappreciate effects of⇌
  - Arousal on sexual behavior⇌
  - Peer pressure on drugs, alcohol⇌
  - Effect of alcohol on smoking, sex, etc⇌
- Most people have *some* awareness and sophistication.⇌
- But still there is a big worry about underestimating these even when (as with hunger) there is a huge amount of past evidence.

↩

Mispredicting Tastes

Addiction/Habit Formation$\rightleftharpoons$

- People may underappreciate the power of addiction. $\rightleftharpoons$
  - Under-investment in good habits (healthy food, exercise, coffee) $\rightleftharpoons$
  - Over-investment in bad habits (tobacco, alcohol)$\rightleftharpoons$
- If non-addicted, hard to: take into account the fact that consistent consumption will mean that you will both feel bad, and develop a stronger desire to keep consuming.$\rightleftharpoons$
  - Under-investment in "good addictions" (e.g., appreciation of classical music, exercise, coffee), and more importantly$\rightleftharpoons$
  - an over-development of bad addictions: tobacco, alcohol, etc.$\rightleftharpoons$
- Once addicted, underappreciating fluctuations and cues:$\rightleftharpoons$
  - binges, and (costly) failed attempts to quit.$\rightleftharpoons$

The role and power of habits may be massively underappreciated by economists ...

$\hookleftarrow$

Mispredicting Tastes

"Nothing so needs reforming as other people's habits."
— Mark Twain

"The diminutive chain of habit is scarcely heavy enough to be felt till it is too strong to be broken."
— Samuel Johnson (simplification from 1800's temperance literature)

"The child is the Father of the Man."
— William Wordsworth

"The second half of a man's life is made up of nothing but the habits he has acquired during the first half."
— Feodor Dostoevsky

"Men's natures are alike; it is their habits that separate them."
— Confucius

"My problem lies in reconciling my gross habits with my net income."
— Errol Flynn

Mispredicting Tastes

Excess wealth-seeking/consumption?⇌
Don't fully appreciate how pleasure from future standard of living will decrease once we become accustomed to that standard of living.⇌

- *IF* consumption is more habit forming than leisure, may suggest that people excessively pursue wealth.⇌
- Thinking lots more money forever will bring us lots more happiness forever. But we'll get used to it.

↩

Mispredicting Tastes

Recently, scale of economic growth as source of well-being (in already-wealthy nations) called into question.$\rightleftharpoons$

- Misprediction might play a role.$\rightleftharpoons$
- This is in part based on better attention to the psychology/sociology of well-being, and part on purported panel and cross sectional empirical evidence that growth does not seem to bring much happiness.$\rightleftharpoons$
- (Research is on-going ... I believe non-adaptive models have exaggerated benefits of growth, but also very skeptical of over-strong claims doubting that money buys happiness.)

$\hookleftarrow$

Mispredicting Tastes

Some reasons that economic growth in nations may not buy much happiness:⇌

- 1. "Classical externalities": E.g., pollution, incessant marketing⇌
- 2. Social/Interpersonal Comparisons⇌
- 3. Intrapersonal comparison/habituation⇌

Note that the first two may call for skepticism about growth without assuming any irrational behavior, since they involve externalities.⇌

- But the third is especially compelling only if we posit some irrationality. If it is true that money doesn't buy happiness (for long), why would rational people work so hard to get more money?

↩ρ

Mispredicting Tastes

Purchase of a Durable Good$\rightleftharpoons$

- Initial excitement will decay over time. $\rightleftharpoons$
  - Especially excitement caused by "sales hype".$\rightleftharpoons$
- Also random fluctuations in excitement
- Fully rational person would take decay into account.

$\hookleftarrow$

Mispredicting Tastes

Projection bias: underestimate decay and fluctuations. Hence:$\rightleftharpoons$

- Firms may choose socially excessive levels of sales hype. $\rightleftharpoons$
- "Artificial durableness": firms create long-term contract (e.g., magazine subscriptions) based on current excitement.$\rightleftharpoons$
- If decay, always too likely to buy.$\rightleftharpoons$
- If excitement exogenous, random, non-decaying.$\rightleftharpoons$
  - Any given period, mistake both directions.$\rightleftharpoons$
  - Buy when most excited $\longrightarrow$ too likely eventually buy

$\hookleftarrow$

Mispredicting Tastes

Hard-to-Reverse Decisions

In difficult-to-reverse decisions:⇌

- Projection bias leads people to underappreciate the degree to which the "hot" feelings will dissipate.⇌
- Too likely to make irreversible decisions. ⇌
  - Getting married in the heat of passion⇌
  - Committing suicide in the depth of depression⇌
  - Sending that e-mail under the grip of rage⇌

↩℘

Mispredicting Tastes

Both normative policy analysis and descriptive political economy:

- Cooling-off periods.

Mandatory "cooling-off period" for sales contracts:

- Customer has option to back out for some period?
- Cost: delays transaction date, prevents trade that is inherently short-term, makes hard for trading partner to rely on trade.
- Benefit if people have projection bias:
  - People less likely stuck with bad purchases.
  - Decreased incentives for sales hype to induce temporary excitement.

↩

Mispredicting Tastes

Revealed Preferences & Revealed Predictions of Preferences$\rightleftharpoons$

- We should not be in the business of condescendingly deciding what people's preferences 'ought' to be.$\rightleftharpoons$
- Within reason, *De gustibus non est disputandum*$\rightleftharpoons$
- But "within reason" matters ... $\rightleftharpoons$
  - Attributing behavior to *functionally* insane 'utility maximization' rather than real mistakes is not humility, or anti-paternalism. It is far less insulting *and* far better science to attribute behavior to human mistakes than to inhuman preferences.

$\hookleftarrow$

Mispredicting Tastes

And (for economists) the mother of all functionally insane preferences:⇌

- **throwing away money**⇌

The best friend of Psychology & Economics in the whole wide world is the same as the best friend in the whole wide world of the rest of Economics: P_____ T_____.⇌

- **Price Theory**⇌

It's more important than even P_____ T_____!⇌

- **Prospect Theory**⇌

How do we reveal potential misprediction of preferences? (Revealing certain types of statistical mistakes is in fact easier than this.)

↩℘

Mispredicting Tastes

Recall addiction example:$\rightleftharpoons$

$$u_t(a_t = y | a_{t-1} = n) = 5$$
$$u_t(a_t = n | a_{t-1} = n) = 0$$

$$u_t(a_t = y | a_{t-1} = y) = -2$$
$$u_t(a_t = n | a_{t-1} = y) = -12\rightleftharpoons$$

If these are consumption utilities, and as part of complete model you tried to look at price effects? Suppose:$\rightleftharpoons$

- Prices $(p_1, p_2, p_3, ..., p_T) = (3, 9, 9, ..., 9)$ generated $(a_1, a_2, a_3, ..., a_T) = (y, y, y, ..., y)\rightleftharpoons$
- Prices $(p_1, p_2, p_3, ..., p_T) = (6, 6, 6, ..., 6)$ generated $(a_1, a_2, a_3, ..., a_T) = (n, n, n, ..., n)\rightleftharpoons$
- Unless person dislikes \$, have revealed misprediction.

$\curvearrowright$

Mispredicting Tastes

I think preference misprediction/uncertainty may be real big-ticket item in positive and (especially) welfare economics.⇌

- Mis-optimizing like some of examples of fluctuating/satiating preferences⇌
- Big life choices? ⇌ How do we have any idea of utility of ⇌
  - Career choice?⇌ (e.g., leaving the farm)?⇌
  - Emigration?⇌
  - Moving away from California?⇌
  - Marriage?⇌
  - "Coming out"?⇌
  - Growing old?⇌
  - **Children?**⇌
  - Habits, habits, habits, habits, and habits.⇌
- Other uncertainties and biases than projection bias.

↩