Disclaimer

- These slides are posted solely to accompany Matthew's lectures at the 2016 Russell Sage Summer Institute in Behavioral Economics.
- They are not edited carefully as stand-alone notes, and are not intended for general circulation.

More Realistic Preferences

Next 3 lectures: improving assumptions about preferences:

- Belief-Based Preferences
- Reference-Dependent Utility, Risk Preferences, News Utility
- Social Preferences

Belief-Based Utility

In psychology for very long, and more recently in formal economics, researchers contemplate preferences that depend directly on beliefs.

- Not $U(c)$, where c is "physical"/"experienced"/ outcomes
- But $U(c, p)$, where p is (current or recent) beliefs about something.
- Care about beliefs not solely for "instrumental" reasons.
- Differs from how beliefs enter choice under expected-utility.
  - In EU: Beliefs over utility, not utility over beliefs
- Linearity of p matters a lot behaviorally
  - EU is weight assigned to different eventualities
  - Not description of states of mind
- Normatively compelling when weighting eventualities

Belief-Based Utility

- Whether food you just ate is *truly* poisonous or unhealthy, will have an important effect on your well-being down the line.⇌
- But your *beliefs*, even if not true, affect your well-being *now*.⇌
  - If *think* food poisonous, you'd be in (very unhappy) distress.⇌
  - If *think* food fattening, guilt or negative anticipatory utility⇌
  - If *think* food not exploitatively produced, be happier⇌
- My prediction: Putting the p in utility is the next big trend.⇌
  - How long have I been making that prediction?⇌
- Utility maximization, much like old.⇌
- But some difference with old-style modeling. ⇌
  - Connection between correct beliefs and happiness?⇌
  - Time-consistency issues⇌requires researchers cope ...⇌
    - a) emotionally, and ⇌
    - b) technologically.

Belief-Based Utility

Examples where well-being and choice importantly influenced by beliefs.

- Ego utility/self-image/identity
- Anticipation and dread.
- Disappointment and pleasant surprises
- Social status and envy etc.
- Health anxiety (and attitudes towards testing)
- Social image—what do others think of us?
- Anger. Grief. Jealousy.
- Sports, movies, novels, news (before vs. after results known).

In each: happiness mediated by our (true or false) beliefs about world.

- My view: **most** preferences are belief-based.
    - Food for (later) thought: includes all choice-set-dependent prefs.
- Obsession with b-b preferences *not* widespread in BE.

Belief-Based Utility

Classical economic issue: $\rightleftharpoons$

- Unemployment. $\rightleftharpoons$
- Classical assumptions $\longrightarrow$ hard to generate big individual cost of unemployment. $\rightleftharpoons$
- Intuition, and (non-decisive) happiness evidence suggests: $\rightleftharpoons$
  - effect on well-being greater than captured by economic models $\rightleftharpoons$

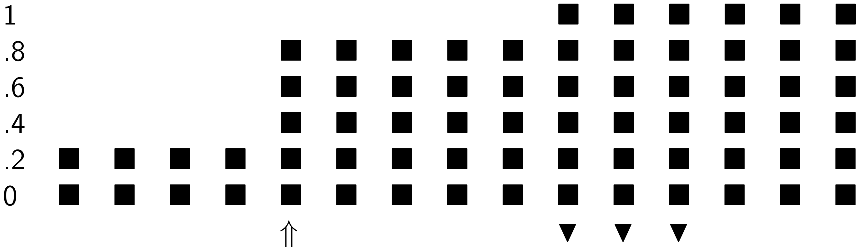
We have not had the formal apparatus to deal with: $\rightleftharpoons$

- Self-image consequences of being unemployed. $\rightleftharpoons$
- Social-image consequences of being unemployed. $\rightleftharpoons$
- Anxiety of becoming unemployed. $\rightleftharpoons$
  - Not merely behavioral (spend less, no vacations, etc.).

Belief-Based Utility

Example to frame thinking: $\rightleftharpoons$

- For a long time, thought probably no 3-day vacation. $\rightleftharpoons$
- Then one day find out that probably will (80%),$\rightleftharpoons$
- and then confirmed as 100% likely when it happens.$\rightleftharpoons$
- Belief evolution: $\rightleftharpoons$



$\looparrowright$

Belief-Based Utility

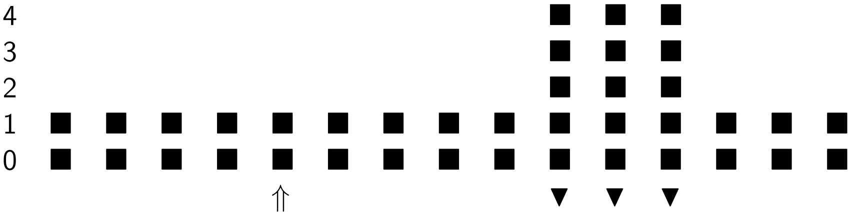
Utility in time?$\rightleftharpoons$

- Will talk about real-time "happiness" without choice.$\rightleftharpoons$

Stay calm!

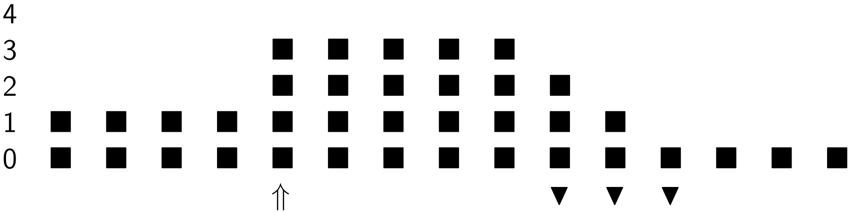
- Will have implications for choice.$\rightleftharpoons$

Utility if above beliefs? $\rightleftharpoons$Could be:$\rightleftharpoons$



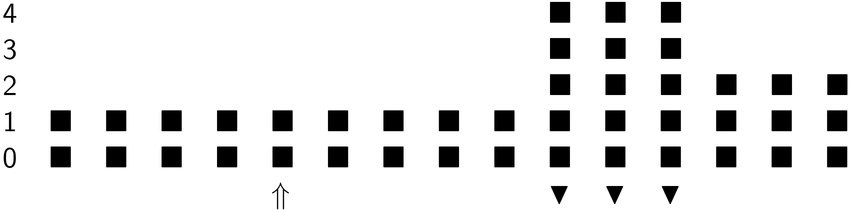
Belief-Based Utility

Or could be:$\rightleftharpoons$



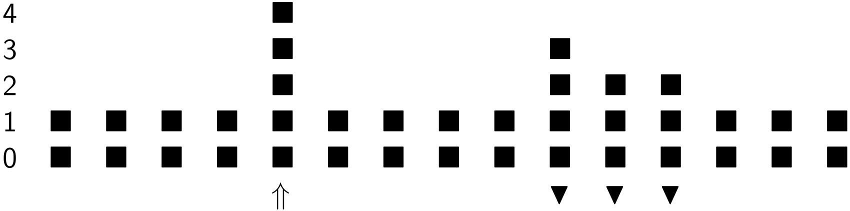
$\rightleftharpoons$

Or could be:$\rightleftharpoons$



Belief-Based Utility

Or could be (my personal vote):$\rightleftharpoons$



$\looparrowright$

So what?

- Why care about the timing or reason for enjoying a vacation?
- Often: We don't. All captured by *utility*(*vacation*).
  - Reduced form probably best for "remembered utility".

But can matter for various reasons. Three are:

- Use direct happiness data iff our theories specify timing of utility.
  - (Not topic these lectures).
- Beliefs/information matter even when behavior is unaffected.
- Affects choice: including time inconsistency, commitment, etc.

Belief-Based Utility

Suppose planning vacation:

- Have anticipatory preferences *for holidaymaking only*.
- \$10,000 Club Medic holiday package, *total* anticipatory utility plus consumption and remembered utility well worth \$10,000.
- But without anticipatory utility, *not* nearly worth it.
- Can/must buy months in advance.

Situation A: All but \$50 is fully refundable if 24 hours in advance.

- What would a fully rational person (as I define it) do?
- She **would/would not** (**cross out one**) buy the package, and then she **would/would not** (**cross out one**) go on the vacation.
    - She **would/would not** (cross out one) buy the package, and then she **would not** go on the vacation.
- And so:
    - She **would not** buy the package, then **would not** go on the vacation.

Belief-Based Utility

If fully rational and have the specified preferences,$\rightleftharpoons$

- You won't sign up under Policy A, $\rightleftharpoons$
  - because you'll cancel, $\rightleftharpoons$*and know you'll cancel*$\rightleftharpoons$
- Won't get anticipatory utility after all.$\rightleftharpoons$

"Fully rational" defined:$\rightleftharpoons$

- Dynamically optimal, ancitipating correctly own conduct.$\rightleftharpoons$
- But not the beliefs that make you happiest.$\rightleftharpoons$
- With b-b preferences, the two are different.$\rightleftharpoons$
  - Remember: not the same under classical EU$\rightleftharpoons$

Situation B: Contract allows no refunds.$\rightleftharpoons$

- What would a person do?$\rightleftharpoons$
- Buy package? Go on vacation?$\rightleftharpoons$
  - She **would** buy the package, and then she **would** go on the vacation.

Belief-Based Utility

Predictions in A&B **do/do not** (**cross out one**) violate classical assumptions about preferences?

- This **does** violate classical axioms/assumptions about preferences.
  - Chose plan "No Buy" from {No Buy, Go, Cancel} in Situation A.
  - "Go" from {No Buy, Go} in Situation B.
  - (And worse off in Situation A)

This **should/should not** (**cross out one**) freak you out?

- This **should not** freak you out.

Violating such an axiom **should/should not** (**cross out one**) thrill you?

- It **should not** thrill you.

Instead be interested in realism, insight, and importance of assumptions.

- (Not sure of realism, importance this example)

Belief-Based Utility

Lessons from this example:

- Generates Choice-Set-Dependent Preferences.
- Correct beliefs vs. happiness-inducing beliefs?
  - If believed would go through with it?
- Welfare effects, choice-set effects, must be handled with care.
- Making predictions:
  - Kőszegi: "personal equilibrium," KR: PPE related formulations.
- "De gustibus" philosophy in its purest and most legitimate form:
  - anticipatory preferences described here are perfectly coherent;
  - we should not rule them out out of principle.

Belief-Based Utility

Extra slides contain model of consumption & savings with anticipatory preferences...

- A big prediction: without commitment, consumption will be increasing over time
  - Intuition?
  - And with commitment, even more increasing.
- Why from anticipating solely his future consumption utility?
  - Why not also from future anticipatory utility?
  - We'll ignore.

Consumption & Savings with Anticipatory Preferences

Setting and Preferences

Yuki will live for 3 periods, has \$$y$ to spend over that time (no interest), seeks to maximize his (undiscounted) lifetime utility $U^1 = u_1 + u_2 + u_3$.

- In period t, "consumption utility" m_t that depends on c_t.
- Also gets utility from anticipating his future consumption utility.
- Why from anticipating solely his future consumption utility?
  - Why not also from future anticipatory utility?
  - We'll ignore.

Belief-Based Utility

Attempt to model this:

$$u_1 = m(c_1) + \varphi[m(c_2) + m(c_3)]$$
$$u_2 = m(c_2) + \varphi[m(c_3)]$$
$$u_3 = m(c_3)$$

- where $\varphi \geq 0$ is relative concern for anticipatory utility.

Question: what is *incoherent* about such preferences?

- u_1 cannot depend on c_2 or c_3. Only **beliefs** about c_2, c_3.

Belief-Based Utility

2nd attempt to model:$\rightleftharpoons$

- $u_1 = m(c_1) + \varphi E_1\{m(c_2) + m(c_3)\}$$\rightleftharpoons$
- $u_2 = m(c_2) + \varphi E_2\{m(c_3)\}$$\rightleftharpoons$
- $u_3 = m(c_3)$$\rightleftharpoons$

where $E_t\{m(c_\tau)\}$ is period-t expectations of period-τ cons. utility.$\rightleftharpoons$

- Would want more complete version of this if there is uncertainty.$\rightleftharpoons$

When beliefs deterministic, shorthand:$\rightleftharpoons$

- $u_1 = m(c_1) + \varphi[m(\tilde{c}_2^1) + m(\tilde{c}_3^1)]$$\rightleftharpoons$
- $u_2 = m(c_2) + \varphi[m(\tilde{c}_3^2)]$$\rightleftharpoons$
- $u_3 = m(c_3)$$\rightleftharpoons$

where $\tilde{c}_\tau^t$ are Yuki's period-t beliefs about period-τ consumption.$\rightleftharpoons$

- What will Yuki do?

Belief-Based Utility

Candidate solution: Yuki solves

$$\text{Max}_{\ c_1, c_2} = m(c_1) + (1+\varphi)m(c_2) + (1+2\varphi)m(y - c_1 - c_2).$$

- E.g., if $m(x) = \ln(x)$, then
  - $c_1^{**} = \frac{1}{3+3\varphi}y$, $c_2^{**} = \frac{1+\varphi}{3+3\varphi}y$, $c_3^{**} = \frac{1+2\varphi}{3+3\varphi}y$
- How do these depend on φ?
  - Respectively decreasing, independent of, and increasing in φ
  - Intuition?
- If $\varphi = 1$, then:
  - $c_1^{**} = \frac{3}{18}y$, $c_2^{**} = \frac{6}{18}y$, $c_3^{**} = \frac{9}{18}y$
  - Match for evidence?

Belief-Based Utility

Is this what Yuki will do?⇌

- Claim: we have under-specified features of the environment.⇌
  - We need to say when Yuki is making (committed) choices.⇌
- Situation 1:⇌
  - Yuki fully rational and can commit, then yes.⇌
- Situation 2:⇌
  - Yuki fully rational and *cannot* commit, then only 1 of the 6 is right.⇌

E.g., if $\varphi = 1$, then⇌

Belief-Based Utility

	Can Commit	Cannot Commit
c_1^*	$\frac{3}{18}y$	$\frac{3}{18}y$
c_2^*	$\frac{6}{18}y$	$\frac{5}{18}y$
c_3^*	$\frac{9}{18}y$	$\frac{10}{18}y$

What is interesting?

- Consumer more period 2 with commitment than without!
- Why does commitment increase period-2 consumption?
  - Because assumed anticipation is over future consumption utility alone (and not future anticipatory utility), happier looking forward to smoothed consumption than back-weighted consumption.
  - But in period 2, this is no longer a consideration.

Belief-Based Utility

Reasons increased consumption profiles besides anticipatory?

- Precautionary savings
- Backward-looking habit formation

Reasons we may rarely see increasing consumption?

- Present bias... consumption smoothing may be self-control problem.
- Because... anticipatory isn't quite right.
  - It is naturally mis-*ID* for its cousin, news utility.
  - Disappointment and surprise utility
  - must balance anticipation with threat of disappointment.
- Reminder: models should own *all* their implications
  - Anticipatory utility makes some strange ones.

Before very different form belief-based utility, anticipatory utility and savings context for a model you may be familiar with.

Belief-Based Utility

Stepping (well) outside the rational framework: $\rightleftharpoons$

- What if Yuki can fool himself into believing lifetime income y is something else?$\rightleftharpoons$
- What might he tell himself?$\rightleftharpoons$
  - Choose to be optimistic to consume anticipation.$\rightleftharpoons$
  - But trades off against induced under-saving.$\rightleftharpoons$
  - Brunnermeier and Parker (2005).$\rightleftharpoons$
- How well-being of such "rationally irrational" beliefs compare to stuck-with-rationality utility?$\rightleftharpoons$
- But ... what if Yuki can tell himself other stories?$\rightleftharpoons$
  - Like that he earns lots of interest rate on his savings?$\rightleftharpoons$
  - Or wonderful afterlife if maximize true lifetime utility.$\rightleftharpoons$
- Fundamental Theorem of Optimal Distortion of Anticipatory Prefs:$\rightleftharpoons$
  - If no restrictions, then choose beliefs to maximize **both** anticipatory preferences and "direct-consumption" utility.

Belief-Based Utility

More generally, models (that many of us have tried) for “motivated” *willful* distortion of beliefs, if not Bayesian (as about to see!) run into problems ... $\rightleftharpoons$

- Psychological realism?$\rightleftharpoons$
- Need a model of what are the limits to distortions.

$\looparrowright$

Ego Utility

Kőszegi (2006a, 2006b) and others: how one feels about oneself.

- We do so here in the context of ego about skill/talent.
- Nascent research: "social preferences" as self & social image:
  - (Related papers: Andreoni on warm glow; Bernheim on conformity; Rabin on moral preferences vs. moral rules; Benabou & Tirole; Grossman; Andreoni & Bernheim; moral wiggle room)
- Status
- We like to think we are:
  - brilliant economists,
  - awesome curlers, and
  - good looking.
- **We like to think we are good people**.

Belief-Based Utility

Different approaches to investigating b-b preferences: Combine assumptions about the psychology, the informational environment, and researchers' methods.
Some approaches:

1. Non-linearity in probabilities, fully-rational/Bayesian
2. "Willful" violations of LIE
3. Self-signaling and forgetting
4. Revealed Altruistic Preferences
    1. Emotional Agents
    2. Responses of Emotional Principals

- Approaches 1 and 3 are most common in recent formal models.
- Approaches 3 and 4 are in fact often formally quite similar.
    - And Approach 2 is a close cousin of these.

Belief-Based Utility

We'll do Approach 1.$\rightleftharpoons$

- Based on Koszegi (2006)$\rightleftharpoons$
- Person will make choices as to whether acquire information.$\rightleftharpoons$
  - Certainly the conceptually easier case.$\rightleftharpoons$
  - Such models of Bayesian self-persuasion pre-date models of Bayesian other-persuasion.$\rightleftharpoons$

Alternative (won't do):$\rightleftharpoons$

- "Self-signaling": By actions or information want to convince future self something different about yourself than now know.$\rightleftharpoons$
  - Reduces to a classical two-person signaling situation, but still a conceptual challenge.$\rightleftharpoons$

Similar dichotomy in social preferences:$\rightleftharpoons$

- Rabin (1995) moral-rules approach non-linear$\rightleftharpoons$
- Recent signaling models.

Belief-Based Utility

There are two types of people in the world:

- Those good at widget-making, and those who are not.

Henrik has utility function $u = w - e + \varphi\sqrt{p}$, with parameter $\varphi \geq 0$,

- where w is his income,
- e is his effort costs of widget-making,
- p is his believed probability that he is good at widget-making.

Notice the embedding:

- If $\varphi = 0$, "classical", purely instrumental prefs.
- If $\varphi > 0$, Henrik has "ego utility".
- (But is it Peemish? What do we have ego about?)

Belief-Based Utility

Notice the non-linearity of $u = w - e + \varphi\sqrt{p}$.

- No reason to expect linearity.
- Henrik is "*information-averse*": concave in $p \longrightarrow$ he prefers to have less information rather than more.
  - Henrik prefers beliefs .7 to 50/50 chance of having beliefs .6 or .8
  - He'd of course prefer beliefs .8 to beliefs .7.
  - But Bayesian approach says cannot choose .8 over .7.
- Classical economics: people weakly prefer more information to less.
  - So information aversion is simplest marker of b-b preferences.

Belief-Based Utility

Suppose Henrik has priors $\overline{p} \in [0,1]$ good at widget-making.

If he works, $e = k > 0$, **and he finds out for sure whether good.**

- If good, $w = 1$;
- if bad, $w = \alpha < 1$.

If no work, $e = 0$, $w = 0$, **and he learns nothing.**

Utility from widget-making: $u = \overline{p}[1 + \varphi\sqrt{1}] + (1 - \overline{p})[\alpha + \varphi\sqrt{0}] - k$.

Utility from non-widget-making: $u = \varphi\sqrt{\overline{p}}$.

- For what values of $\overline{p}, \varphi, \alpha, k$, will Henrik work?

If $\varphi = 0$ — classical, non-belief-dependent preferences — then

- If k is low, work for sure.
- more likely to work when $\overline{p}$ is higher (since $\alpha < 1$).

Belief-Based Utility

For $k = .001$, $\alpha = \frac{31}{48,}$, and the combinations of $\overline{p}$, φ, will Henrik work?

$\overline{p}$ and $\varphi =$	$\varphi = 0$	$\varphi = 3$	$\varphi \to \infty$
$\overline{p} = 0$	Yes	Yes	Yes
$\overline{p} = \frac{1}{9}$	Yes	Yes	No
$\overline{p} = \frac{1}{4}$	Yes	No	No
$\overline{p} \to 1$	Yes	Yes	?

Fascinating things about this table?

- Despite $k \approx 0$, Henrik sometimes avoids work.
    - "Paying" to avoid information.
- Fixing any $\overline{p} \neq 0$, $\overline{p} \neq 1$, Henrik will not work as $\varphi \to \infty$.
- Fixing any φ, $\overline{p} \to 1$ or 0, Henrik will work iff exp(wage) $>$ effort.
- Non-monotonicity in $\overline{p}$.

Belief-Based Utility

If $\varphi > 0$ moderate, can get: when $\overline{p}$ goes from very low to moderate, Henrik may go from working to not working.

- Why?

Lesson emphasized by Koszegi (2006): can get negative correlation between confidence about your talent and whether undertake task.

- Protecting ego, not likely without b-b prefs

Belief-Based Utility

And now for a tale of love and widgets ... Recall:

- If $\overline{p} = \frac{1}{4}$, $k = 0, \varphi = 10,000, \alpha = 0$ Henrik no work.

The twist: Suppose that

- Bjorn loves Henrik and wishes him nothing but the best,
- knows whether Henrik is a good widget-maker or not.
- Henrik knows Bjorn loves him and is aware of Henrik's widget skills.
- Common knowledge that Bjorn could communicate with Henrik
- If and when Bjorn says something, Henrik can take job or not

How does this tale of love and widgets end?

- Quite sadly, on average.
    - Henrik will find out whether or not he is a good widget maker.
    - Which makes him feel worse on average.
    - The logic is that of "emotional agency": credible equilibria in revealing information to people with belief-based preferences.

Forget Bjorn.

- Suppose that Henrik instead has preferences: $u = w - e + \varphi p^2$.
  - Henrik is information-loving.
- Harder to see over-gathering of information than under-gathering (since value of information always non-negative).
- Henrik may take job as widget maker in situations where wouldn't for "classical" reasons.

Belief-Based Utility

Now suppose instead that Henrik has preferences: $u = w - e + \varphi p$.

If $\varphi = 0$, what would we observe?

- Henrik obtains information for solely for instrumental reasons.

If $\varphi > 0$, what would we observe?

- Henrik obtains information solely for instrumental reasons.

If $\varphi < 0$, what would we observe?

- Henrik obtains information solely for instrumental reasons.

LIE $\rightarrow$ Henrik's *observed* taste for information entirely instrumental.

- When linear, belief-based preferences can matter a lot for well-being even when no ID in behavior.

How would Bjorn behave?

- Emotional agents can reveal b-b preferences even when linear.
  - "Altruistic revealed preference"

Belief-Based Utility

Conclusion:$\rightleftharpoons$

- Recall: similar themes for self-signaling models.$\rightleftharpoons$
- Recall: one person in room thinks humans all about b-b preferences.$\rightleftharpoons$
- And that they will be big thing in improved preference models.$\rightleftharpoons$

Now turn to Reference Dependence$\rightleftharpoons$

- Will come back to b-b preferences with a vengeance.$\rightleftharpoons$
- Aside: Ref points everywhere... including ego utility, etc.

$\looparrowright$

Reference-Dependent, News-Utility, and Risk Preferences

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.

- Feelings and choice are reference-dependent:
- $u(c_t; r_t)$, not $u(c_t)$, where r_t is some reference level.

Prospect-Theoryish Stuff

Kahneman and Tversky (1979, p. 277) stress that salience of changes from reference points is a basic aspect of human nature:

- "Our perceptual apparatus is attuned to the evaluation of changes or differences rather than to the evaluation of absolute magnitudes ... The same principle applies to non-sensory attributes such as health, prestige, and wealth."

Not just people, but amoeba and physical systems.

- Role of contrast/comparison poster child for things economists should use to explain things, not something to be explained.

Two features of reference dependence emphasized by Kahneman and Tversky (1979) and others:

- Loss Aversion
- Diminishing Sensitivity

Prospect-Theoryish Stuff

1. Loss Aversion

People dislike losses (2 times?) more than they like comparable-size gains.

- E.g.: Vast majority turn down 50/50 lose \$600, gain \$700 bet
  - **Not** because of DMU(W).
  - Strongest such aversion involve mix gains and losses.
  - Many folk, myself included, guilty of sloppy wording, calling all modest-scale risk aversion, even when only gains, "loss aversion".
- Monetary risk attitudes primary area in which LA explored.
- Other domains where LA important:
  - Moral/Fairness, Hippocratic Oath, Wage Cuts
  - [Endowment Effect/SQ Bias in riskless trades]
  - Disposition effects, in investments and houses.
  - Aversion to (nominal) wage and consumption declines.
  - Income-targeting

2. Diminishing Sensitivity$_{\rightleftharpoons}$

- People pay less attention to incremental differences when changes are further away from reference point.$_{\rightleftharpoons}$
- Prefer \$420 for sure or 50/50 chance at \$900? $_{\rightleftharpoons}$
- **Prefer losing \$420 for sure or 50/50 chance to lose \$900?**$_{\rightleftharpoons}$

Reflects big and general fact about human psychology:$_{\rightleftharpoons}$

- We think in terms of proportions.

$\looparrowright$

Prospect-Theoryish Stuff

Change subject: in following pairs, which "feel" like a bigger difference?

visually 101 ft. away vs. 100 ft. away		1 ft. v. 0 ft.
gaining $101 v. gaining $100		gaining $1 v. gaining $0
losing $101 v. losing $100	vs.	losing $1 v. losing $0
losing $101 v. losing $100		losing $2 v. losing $1
gratification 100 days from now v. 101 days		gratification 0 days v. 1 day
saving $10 on $1,000 item		saving $10 on $20 item
carrying a suitcase 21 blocks v. 20 blocks		2 blocks v. 1 block
19% chance v. 18% chance		1% chance v. 0%
19% chance v. 18% chance		100% chance v. 99%
17 heads/13 tails v. 16 heads/14 tails		4 heads/0 tails v. 3 h/1 t

Prospect-Theoryish Stuff

Proportional thinking via diminishing sensitivity in the value function.$\rightleftharpoons$

- But DS and proportional thinking the same?$\rightleftharpoons$
- Distance vs. Range?$\rightleftharpoons$
  - DS: with $r = 0$, 1 vs. 2 in feels bigger than 91 vs. 92 in $\{0, 1, 2, 91, 92\}$$\rightleftharpoons$
  - RT: 1 vs. 2 in $\{0, 1, 2, 91, 92\}$ feels smaller than 1 vs. 2 in $\{0, 1, 2\}$$\rightleftharpoons$
- Bushong-Rabin-Schwartzstein think *some* DS really propo-thinking.$\rightleftharpoons$
  - But some DS really DS$\rightleftharpoons$
- Two more issues: $\rightleftharpoons$
  - Is this really rational "experienced-utility maximization"?$\rightleftharpoons$
  - Seriously contradicts some intuitions/popular recent models that people pay too *much* attention to increments over bigger ranges.

$\looparrowright$

Modeling Reference-Dependent Utility

- Fixing r, rational people will maximize $u(c|r)$.
    - What is r?
    - Why is it fixed?
    - Moving on...
- Reference dependence in $u(\cdot|\cdot)$ fully rational?
    - In principle: yes.
    - People feel pleasure and pain from gains and losses.
    - But wednesday two major ways that the influence of reference-dependent preferences not fully rational
    - Both lead too much attention on gains and losses

Prospect-Theoryish Stuff

Traditional focus: solely on departures from the reference point. $\rightleftharpoons$

- Not how the reference point is formed, $\rightleftharpoons$
- **Nor how the reference point affects utility**.$\rightleftharpoons$

That is, analyze choice in terms of "value-function".$\rightleftharpoons$

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

Wildly insufficient for integrating this into economics.$\rightleftharpoons$

- But I start by framing evidence in terms of that approach$\rightleftharpoons$
- Before returning to the full model$\rightleftharpoons$
- Including how people also care about absolute levels.

$\looparrowright$

Prospect-Theoryish Stuff

Reference Dependence and Risk: Round 1:

- Assume people maximize expected value of $u(c|r)$.
- I tend to ignore non-linear probability weights.
  - Mostly simplicity Partly skepticism.
  - Most traumatizing picture in all of economics!
  - Overweight low probabilities...except when ignore
- Others also have downplayed.
- But it's making a comeback...
  - Largely because it should (I was/we were wrong)
  - Employs 2^{nd}-most traumatic picture.
- Ignoring non-linear probability weights: "reference-dependent Von Neumann-Morgenstern preferences" (RD-VNM prefs):
  - For each r, people maximize $\int_c p(c)u(c|r)$.

Prospect-Theoryish Stuff

Assumptions A0-A4 (K&T (1979), formalized by BMR (1999)):

- A0. $v(x)$ continuous, differentiable for all $x \neq 0$, $v(0) = 0$.
- A1. $v(x)$ is strictly increasing.
- A2. If $y > x > 0$, then $v(y) + v(-y) < v(x) + v(-x)$.
- A3. $v''(x) < 0$ for $x > 0$ and $v''(x) > 0$ for $x < 0$.
- A4. From $x > 0$ direction, $\lim_{x \to 0} \frac{v'(-x)}{v'(x)} \equiv L > 1$.

Implications for Risk Attitudes? (Assume r = current wealth)

- 1. Person will turn down any 50/50 lose \$X/gain \$X bets.
  - Implied by A2.
- 2. Is risk averse among bets involving only gains
  - Implied directly by A3.
- 3. Is risk-*loving* among bets involving only losses
  - Also implied directly by A3.
- 4. Is "first-order risk-averse."
  - Implied by A4.

Prospect-Theoryish Stuff

- Implications 1 & 2 true of EU/DMU(W) model.$\rightleftharpoons$
- Implication 3 famously inconsistent with EU/DMU(W).$\rightleftharpoons$
  - But generates the question:$\rightleftharpoons$
  - Huh? Insurance?$\rightleftharpoons$
- Implication 4: Folk wisdom that first-order risk aversion is $\rightleftharpoons$
  - *inconsistent with* EU/DMU(W).$\rightleftharpoons$
  - But hugely important, not just a limit result.

$\looparrowright$

Prospect-Theoryish Stuff

DMU(W) does not help explain Modest-Scale Risk Aversion

- Virtually none of aversion to risks involving little chance of gaining or losing more than (say) $10,000 can be explained.

What mean by EU/DMU(W) "does not **help**"?

- Not that it is missing factors.
- Rather: for modest stakes, capturing **very** little.
- **Not** that EU/DMU(W) model wrong ... all models are wrong.
  - For **this** question, it has virtually zero explanatory power.
  - It predicts virtual risk neutrality.
  - Just because all useful models are false (because all models are false), doesn't mean that all false models are useful.

EU/DMU(W) *does* help for large stakes... main story.

Prospect-Theoryish Stuff

Four final comments (for now) on risk prefs:

- 1. People *are* in general averse to modest-scale risks.
  - And sometimes risk-loving. Rarely risk-neutral
- 2. Modest-scale risk aversion is an important topic for economics.
  - Extended warranties, aversion to deductibles, fixed-rate plans, etc. are market-driving, institution-driving behaviors.
- 3. Risk aversion neither simplifying assumption nor done for generality.
  - It is one of the most results-driving assumptions in all of economics
  - Ubiquitously added as *necessary* "complication"
- 4. We don't want to get rid of the $u''(w) < 0$ assumption.
  - Among most intuitive, compelling, important assumptions we make!

Prospect-Theoryish Stuff

Four final comments (for now) on risk prefs:

1. People *are* in general averse to modest-scale risks.

- And sometimes risk-loving.
- Rarely risk-neutral

Example from Sydnor (2010) on housing insurance:

Deductible	Premium	Choice
\$1000	\$504	
\$500	\$588	
\$250	\$661	X
\$100	\$773	

- With average claim rate (3.7%), paid \$157 for an expected savings of \$28 — very risk-averse!

Prospect-Theoryish Stuff

2. Modest-scale risk aversion is an important topic for economics.
   - Extended warranties, aversion to deductibles, fixed-rate plans, etc. are market-driving, institution-driving behaviors.
   - Dismissal of lab evidence as only involving small stakes has reached a whole new aesthetic height in this area:
     - "...but would they be risk averse for small stakes for *large* stakes?"

3. Risk aversion neither simplifying assumption nor done for generality.
   - It is one of the most results-driving assumptions in all of economics
   - Ubiquitously added as *necessary* "complication"
   - Shortcomings current model are sufficiently severe — and sufficiently solvable — that it is worthwhile for economists to investigate additional determinants of risk attitudes.

4. We don't want to get rid of the $u''(w) < 0$ assumption.

- Among most intuitive, compelling, important assumptions we make!
- Far more well-being out of the first \$5 million you spend in life than the last \$5 million.

Prospect-Theoryish Stuff

Aside: there are many measures of risk attitudes.

- Most common measure of risk aversion used is the coefficient of relative risk aversion, $\frac{-wu''(w)}{u'(w)}$, where w is a measure of "wealth".
  - Macroeconomists measure CRRA on huge stakes as somewhere in the single digits or low double digits.
  - Experimentalists measure CRRA on small stakes in single digits.
- DMU(W) pretty well calibrated after all?
- Problem: using different measures of wealth.
  - Which is to say, different definitions.
  - "Wealth" is defined as earnings from the experiment alone.
  - Other papers use monthly income, lifetime wealth, or current assets.
- If same definition used
  - 5 orders of magnitude differences.

Integrating Prospect Theory ...

Kőszegi and Rabin (2006, 2007, 2009) clarifies, modifies, and extends PT with the goal of making the theory of reference-dependent preferences more conducive to general application within economics.

- The theory came out of years of teaching without a satisfying, consistent, and applicable model of reference dependence.

Grand Conceit: Because we (jointly) do two things, Derive reference point from preferences and environment, and derive gain-loss utility from reference points and consumption utility,

- aim at universal formula for constructing reference-dependent preferences from existing (non-reference-dependent) models.
- Not quite there, but many contexts close to improved predictions with zero degrees of freedom added to the models:

Prospect-Theoryish Stuff

First premise: **People don't care solely about gains and losses.**$\rightleftharpoons$

- People don't just react to the sensation of gaining or losing a mug; they care whether they have a mug to drink from.$\rightleftharpoons$
- Don't just enjoy sensation of \$-gain: they like more to spend.$\rightleftharpoons$

Certainly worse for economics to ignore this classical consumption utility then ignoring gain-loss utility.$\rightleftharpoons$

- Integrate "gain-loss utility" component of prospect theory with "consumption utility" studied by economists
  - in fact relate the two strongly.$\rightleftharpoons$
- Prospect theory best interpreted as silent on consumption utility, rather than saying it doesn't exist.$\rightleftharpoons$
- But the silence a fundamental barrier in many applications.

$\looparrowright$

Prospect-Theoryish Stuff

Combine Absolute & Contrast Utilities.

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

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

where

- $m(c)$ is "consumption utility", and
- $n(c - r|r)$ is "gain-loss utility".

So far, this decomposition is just an interpretation in our heads.

- Substance to this breakdown given by further structure.
- Actually, dimension-by-dimension.
  - Will suppress dimension notation

Prospect-Theoryish Stuff

We'll assume that $n(0|r) = 0$ for all r, and $n(c - r|r)$ obeys A0 to A4.

- Note that $u(c|c) \equiv m(c)$.
- $m(c)$ is how you feel about outcome that matches reference point.
- Utility matches classical in settings where consume at reference levels!

Prospect-Theoryish Stuff

What is odd about assuming $n(\cdot|r)$ meets A0-A4?

- Behavioral evidence on shape $u(c|r)$, not $n(\cdot|r)$.

Claim: Often okay to assume $m(c)$ is $\approx$ linear, when $c - r$ is small.

- When $m(c)$ is not linear for even small changes ...
  - we predict in fact prospect theory v-function will not hold.
  - we are happy with that prediction

In fact, wrong framing...

Prospect-Theoryish Stuff

Reflecting common intuition, Kőszegi and Rabin (2006) propose:

- gain-loss utility is proportional to changes in consumption utility.

Specifies that $n(\cdot)$ is tightly related to $m(\cdot)$:

- How people feel about gaining or losing the mug is related to how they feel about having the mug to drink from.

Specifically:

- $n(c - r|r) \equiv \mu(m(c) - m(r))$, where $\mu(x)$ is a "universal gain-loss function" meeting properties A0-A4.
- $m(\cdot)$ as the argument helps us address the shape-of-$m(\cdot)$-based exceptions to A0-A4.
- **Big** prediction of including the model: attitudes towards uncertainty at different wealth levels proportional to marginal utility at those wealth levels.

Prospect-Theoryish Stuff

Although limited by the separability assumptions, by a need to specify relevant dimensions, and cardinal interpretation of $m(\cdot)$, this comes close to saying:$\rightleftharpoons$

- Once you specify the reference point, we can propose a zero-degrees-of-freedom alternative prediction to classical reference-free models.$\rightleftharpoons$
- Include news utility, implications for things like precautionary savings.

$\looparrowright$

Stochastic Reference Points

Koszegi & Rabin (2006) also introduce a new issue (and an alternative approach to that taken by others) in modeling reference dependence.

- Clear once we consider beliefs to be the reference point, need to know what happens with a stochastic reference point.
- If a person is endowed with a lottery? What is the reference point?
- If reference point is probability measure G over $\mathbb{R}$, and consumption drawn according to probability measure F, then utility

$$U(F|G) = \int \int u(c|r) dF(c) dG(r).$$

Prospect-Theoryish Stuff

E.g. if endowment is $\frac{1}{3}/\frac{1}{3}/\frac{1}{3}$ lottery \$0, \$50, or \$100, then getting \$50 assessed as loss relative to \$100, gain relative to \$0, and wash relative to \$50:

$$U(\$50|lottery) = \frac{1}{3}u(50|0) + \frac{1}{3}u(50|50) + \frac{1}{3}u(50|100).$$

This model of "mixed feelings" crucial to some of the risk-preferences stuff, less crucial in other contexts.

- But very little evidence. It matches a strong intuition, and some recent evidence. E.g., Sprenger (2010)

Older asserted evidence: when endowed with lotteries.

- Selling price for a lottery is its expected value.
- Takes work to see why this supports stochastic reference points.
  - But it does.

Prospect-Theoryish Stuff

"Mixed feelings" implicitly contradicts common decision-theoretic cousins to this approach.$_{\rightleftharpoons}$

- Those models assume implicitly that the reference point for stochastic outcome is its (recursive) certainty equivalent.$_{\rightleftharpoons}$
- Dynamic-consistency, non-cycling, "resulting-property axioms" may force no-mixed-feelings-allowed property.$_{\rightleftharpoons}$
- But no psychological logic suggesting must be the case.

$\looparrowright$

What is the Reference Point?

- Literature historically fluctuates between imprecision and status quo.
- Typical approach to this issue is in Tversky and Kahneman (1991):

"We assume that the decision maker has a definite reference state X, and we investigate its impact on the choice between options. The question of the origin and 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."

Prospect-Theoryish Stuff

So what is the reference point, already?

- Usually assumed to be status quo,
- or recent consumption
- but often much vaguer.

"What is the reference point?" both example of and metaphor for BE battles.

- Does BE tell you a new reference point in every new situation?
- Response to **clearly true** but underspecified, underformalized theories is to more carefully specify and formalize, not dismiss.
- But worry about limited applicability, limited discipline was legit.

Prospect-Theoryish Stuff

So what is the reference point? Candidates:

1. "Status quo ante"
2. Recent consumption or endowments
3. Aspirations/goals
4. Your neighbors
5. **Expectations**

Discuss later why 1 and 2 might really be 5.

- But I think 3 is definitely about 5 many cases.
    - But see Pope, Simonsohn, etc.
- And 4, comparing ourselves to neighbors, may be very real ...
    - but in a lot of data confounded big time with 5.

Prospect-Theoryish Stuff

The Reference Point as Expectations

- Assume reference point is the probabilistic beliefs about outcomes ("expectations" in shorthand)
- Hence the idea of stochastic reference points.
- Close the model by assuming beliefs obey "rational expectations".

Expectations do not determine the reference point completely.

- Example of an exception: sensory experiences, such as changes in brightness and temperature.

Prospect-Theoryish Stuff

But I would argue expectations are predominant.

- Raises all sorts of issues and complications.
- But we argue that some of the 'complications' are exactly right — we can reconcile things that are variously
  - inconsistent with classical EU but consistent with classical PT,
  - inconsistent with classical PT but consistent with classical EU, and
  - inconsistent with both.

Prospect-Theoryish Stuff

"Disappointment aversion" by Bell, Loomes and Sugden, others, then axiomitized by Gul (1992) and others, pre-cursor of this approach.

- One theme of B-B approach: timing is crucial.
  - Delays in realizations of choices matter a lot, and would need to be added to language of the choice-set-based theories.
  - But how long a person expected the choice set before making a committed choice also matters a lot.
- KR model has degree of freedom, about how quickly reference point adjusts to new expectations.
  - But research exploring ongoing.
- Central premise: different to choose from $\{(.5, -\$250; .5, \$0), -\$100\}$ when
  - a moment ago you thought your choice set was probably $\{\$0\}$ versus
  - if you've known this was your choice set for quite some time.

Prospect-Theoryish Stuff

Evidence + Intuitions: And great classic experiment, oversimplified:

- Medvec *et al* (1995): as judged by photos of heads (above the neck), who's happier? Gold medalists? Silver medalists? Bronze medalists?
  - Obvious econometric issues ... hardly exogenous what people expect and what they achieve.
  - Replication by Mcgraw, Mellers, and Tetlock (2005)
- Unexpectedly losing $50, or a mug you expected to keep, are losses.
- Gaining $50 unexpectedly or getting mug unexpectedly are gains.
  - *But* spending $50 you planned to spend?
  - Losing a mug you planned to sell?
  - Getting $50 allowance you get every week?
  - Getting delivery of mug long anticipated?

Prospect-Theoryish Stuff

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

- Ignoring is **not** an answer!$\rightleftharpoons$
  - Stop claiming your model fits by ignoring when it doesn't!$\rightleftharpoons$
- Answer 1: Most evidence on surprises: status quo = expectations.$\rightleftharpoons$
  - (rationally or psychologically) expect to maintain the status quo.$\rightleftharpoons$
  - E.g., owning a mug often connotes to us that we will be in possession of it tomorrow.$\rightleftharpoons$ (this is cheating)$\rightleftharpoons$
- Answer 2: In fact, often wording obscures things.$\rightleftharpoons$
  - E.g., wage cuts are not decreases in wealth ... they are (surprise) decreases in the rate of increase of wealth!$\rightleftharpoons$
  - How does an unexpectedly small wage increase feel?

$\looparrowright$

Shopping! ⇌

- Now: seemingly important (but largely untested) predictions about a pretty core topic in economics, consumer demand. ⇌
    - Simple example of KR (2006); ⇌
    - market analysis in HK papers, Rosato. ⇌
- "Reference-price effects" in marketing literature ⇌
    - Little emphasized seems among BE-ers studying loss aversion. ⇌
- Two dimensions, shoes and money: ⇌
    - $m(c) = 16 \cdot \textit{pairshoes} + \$$. ⇌
    - "Intrinsic" value for shoes is \$16. ⇌
    - μ is two-piece linear, with $\mu'_+(0) = 1, \mu'_-(0) = \lambda$.

↬

Prospect-Theoryish Stuff

Analysis with parameter values $\eta = 1, \lambda = 3.$$\rightleftharpoons$

- Buying at price p when fully expect to face price p is the optimal consistent plan iff $p \leq 16$$\rightleftharpoons$
- **When no uncertainty WTP is "intrinsic" value for shoes.**

Prospect-Theoryish Stuff

Things change dramatically when uncertainty over prices.

- Suppose beliefs: $prob(p = p_H) = prob(p = p_L) = .5.$

Some examples, with PPE behavior and utility.

- If $(p_L, p_H) = (18, 18)$, Never Buy, $U = 0$
- If $(p_L, p_H) = (14, 14)$, Always Buy, $U = 2$
- If $(p_L, p_H) = (2, 14)$, Always Buy, $U = 2$
- If $(p_L, p_H) = (14, 95)$, Never Buy, $U = 0$
- If $(p_L, p_H) = (4, 18)$, Always Buy, $U = -1.5$
- If $(p_L, p_H) = (1, 18)$, Buy iff $p = p_L$, $U = -1$
- If $(p_L, p_H) = (4, 95)$, Buy iff $p = p_L$, $U = 1$

Prospect-Theoryish Stuff

- If $(p_L, p_H) = (18, 18)$, Never Buy, $U = 0$
- If $(p_L, p_H) = (14, 14)$, Always Buy, $U = 2$
- If $(p_L, p_H) = (2, 14)$, Always Buy, $U = 2$
- If $(p_L, p_H) = (14, 95)$, Never Buy, $U = 0$
- If $(p_L, p_H) = (4, 18)$, Always Buy, $U = -1.5$
- If $(p_L, p_H) = (1, 18)$, Buy iff $p = p_L$, $U = -1$
- If $(p_L, p_H) = (4, 95)$, Buy iff $p = p_L$, $U = 1$

Do you buy at \$18? At \$14?

- "environmental-endogeneity" principle:
  - price distribution influences your WTP.
- But theory puts limits on this (as function of λ, η)
- No matter price-distribution:
  - Never buy when $p > 32$, always buy when $p < 8$
- Agenda: Replace "context effects" with:
  - more fully specified utility models that endogenize and fully specify how and why a context/environment influences choice.

Prospect-Theoryish Stuff

- If $(p_L, p_H) = (18, 18)$, Never Buy, $U = 0$
- If $(p_L, p_H) = (4, 18)$, Always Buy, $U = -1.5$
- If $(p_L, p_H) = (1, 18)$, Buy iff $p = p_L$, $U = -1$
- If $(p_L, p_H) = (4, 95)$, Buy iff $p = p_L$, $U = 1$

Attachment effect:

- if likely price is low, think likely you'll return home with the shoes.
- Increases your WTP even when high price, to avoid loss.

Comparison effect:

- Lowering the low price decreases WTP at high prices—high price feels like a worse loss.

Prospect-Theoryish Stuff

- If $(p_L, p_H) = (18, 18)$, PPE is Never Buy.
- If $(p_L, p_H) = (4, 18)$, PPE is Always Buy.
- If $(p_L, p_H) = (1, 18)$, PPE is Buy iff $p = p_L$.

Two effects sometimes go opposite directions

- No apologies; such is the stuff of classical consumer theory ....
- But notice "quasi-violation" of law of demand:
  - *known* lowering of prices may decrease quantity demanded.
  - Will never happen fixing expectations
  - possibility comes from "raising expectations" for a good deal.
- Requires stochastic prices,
  - no good sense for empirical/calibrational relevance.

Prospect-Theoryish Stuff

- If $(p_L, p_H) = (18, 18)$, PPE is Never Buy.
  - $U = 0$
- If $(p_L, p_H) = (4, 18)$, PPE is Always Buy.
  - $U = -1.5$
- Utility can be lower in PPE than never buying.
  - Sometimes paying more than thought you might, or
  - Sometimes not getting the good you thought you might
  - might outweigh consumption utility of sometimes getting the good at a low price.
- PE consistency constraint here is in force:
  - might want commit not purchase, can't resist low prices.

Punchlines of this approach:$\rightleftharpoons$

- Departures from classical predictions iff surprise or uncertainty.$\rightleftharpoons$
- when uncertainty, environment influences (induced) preferences!$\rightleftharpoons$
- Some desirable plans not consistent ...

$\looparrowright$

Prospect-Theoryish Stuff

Reinterpretation of the endowment-effect evidence.$\rightleftharpoons$

- The endowment effect is commonly explained by assuming that the reference point of status quo$\rightleftharpoons$
- ... sellers keep the mug$\rightleftharpoons$ buyers keep their money.$\rightleftharpoons$
- Experimental protocol *arguably* induce parties to expect to keep endowment.

$\looparrowright$

Prospect-Theoryish Stuff

More natural expectations reverse endowment effect?

- In "real world," buyers expect/hope to part with money, get good. Sellers expect/hope to lose good, get money.
- One interpretation of Plott and Zeiler (2006) and of List (2006) is that traders there have expectations of trade.
- More recent experimental evidence:
    - Ericson and Fuster (QJE): supportive that expectations play a role.
    - Heffetz and List (JEEA), similar experiment, *and others*, reject role of expectations

Prospect-Theoryish Stuff

Risk Preferences Revisited

- Some in KR (2006), main topic of KR (2007), and see KR (2009), including/especially on-line Appendix.

Main theme:

- Replicates KT's Prospect Theory in some *identified* contexts.
- And replicates classical DMU(W) in other *identified* contexts.
- Often corresponds to neither PT nor DMU(W).
    - *Reflects intuitions of PT, but makes very different predictions than status-quo-based PT literally does.*

Prospect-Theoryish Stuff

Both "foes" *and* "friends" over the years seen demand for insurance as inconsistent with convexity-in-losses feature of prospect theory.$\rightleftharpoons$

- KT have explained with probability weighting.$\rightleftharpoons$

Certainly an aspect of risk preferences that is important.$\rightleftharpoons$

- And some recent papers argue it *is* probability weighting$\rightleftharpoons$
  - See, e.g., O'Donoghue et al$\rightleftharpoons$
- But I also don't think it can explain some of the taste for insurance, especially if it has to do battle with DS in losses.$\rightleftharpoons$
- I doubt some of the identification

$\looparrowright$

Prospect-Theoryish Stuff

1. Large-stake risks → classical DMU(W)⇌
   - Depends on curvature and whether large stakes dominated by DMU(W), depends on severity of both DS and concavity of m.⇌
2. Surprise, modest-scale risks → KT's Prospect Theory ⇌
   - First-order risk aversion to 'mixed risks'⇌
   - Risk averse in gains, risk loving in losses.⇌
3. Expected, modest-scale risks w/o commitment → heavy risk aversion⇌
   - With commitment, even possible to reject dominant lotteries.⇌
     - But improved approach of KR (2009) no longer predicts it?⇌
   - Not risk-loving in expected "losses" — because in our model sensation of losses arises only from worse-than-expected decreases.

Prospect-Theoryish Stuff

4. In various ways, people become less risk averse when they expect risk. $\rightleftharpoons$
   - *But:* $\rightleftharpoons$
5. Expecting risk makes you very unhappy. $\rightleftharpoons$
   - "Equilibrium risk" is always bad: the distribution of gains is identical to the distribution of losses. $\rightleftharpoons$
   - And when can buy insurance prospectively (which is the only type of insurance...), become very risk averse. $\rightleftharpoons$

Prospect-Theoryish Stuff

All this talk of money raises a pretty big conceptual point.$\rightleftharpoons$

- KR (2006) follow others in obscuring, cheating, and fudging on:$\rightleftharpoons$

Why do people have utility over money at all?$\rightleftharpoons$

- In classical model, money only valued for consumption.$\rightleftharpoons$
- Legitimate object of analysis as the indirect utility function.$\rightleftharpoons$
- But not current consumption, which KR (2006, 2007) treat it as.

$\looparrowright$

The psychology of disappointment and elation is about news.

- Not just contemporaneous consumption relative to expectations.
- But about changes in beliefs about future consumption.
- Money is ... news about future consumption.

News Utility: KR (2009): reference point is still recent expectations.

- But: Rather than comparing outcomes to beliefs, people compare new beliefs to old beliefs.
- Although the full implications are far more complicated, the basic intuition for what this says about money is very simple:
- Money is future consumption so unexpected gains and losses of money are news about future consumption.

Prospect-Theoryish Stuff

New parameter: $\gamma \equiv$ the degree to which you care about "prospective gain-loss utility" compared to "contemporaneous gain-loss utility".

- Do revisions in beliefs before an outcome occur cause comparable sensations of gain and loss as surprises at time of outcomes?

Prior analysis all implicitly assume $\gamma = 0$.

- Depending on the context, analysis may be misleading.

Prospect-Theoryish Stuff

Not anticipatory utility a la Caplin-Leahy, Kőszegi, Brunnermeier-Parker.

- Our prospective gain-loss utility is about utility from *changes* in expectations of future consumption.

Difference is going to matter *a lot*. And further:

- Because people didn't have news-utility interpretation, obvious confound in evidence with anticipatory utility.
- Ceteris paribus — fixing this morning's expectations — obvious correlation with how optimistic you are about the future and how much more optimistic you have become about the future.

Prospect-Theoryish Stuff

Happiness (especially rich people) in September and October 2008?$\rightleftharpoons$

- Moments in world history where rich folk less happy than poor?$\rightleftharpoons$
  - 1789 France, $\rightleftharpoons$
  - 1979 Iran?$\rightleftharpoons$
  - 2008 "developed world"
- But 2008 rich still had more comfortable future than 2008 poor.$\rightleftharpoons$
  - And surely knew it! $\rightleftharpoons$
  - (1789 rich French folk had different future?)$\rightleftharpoons$
  - Not anticipatory utility, but prospective news utility.$\rightleftharpoons$
- Final twist: $\rightleftharpoons$
  - Perhaps rational inference ... crisis even worse *news* for poor???$\rightleftharpoons$
  - Utility: changes in beliefs ... not whether those beliefs rational.

Prospect-Theoryish Stuff

Aside:$\rightleftharpoons$

- In models of anticipatory utility, anticipated high consumption in a given period enters into the utility in all prior periods.$\rightleftharpoons$
- "Multiple-counting" consumption?$\rightleftharpoons$
    - Intrinsic prediction increasing consumption$\rightleftharpoons$
- News utility does not multi-count *when beliefs stable*.

$\hookrightarrow$

Wealth and Consumption in Intertemporal Choice

- In more detail in K&R (2009).
- See also Pagel (2013)
- Consumer allocates consumption between periods 1 and 2.
- Can't commit ???
- Intertemporal budget constraint is $c_1 + c_2 = W$.
- Consumption utility each period is $m(c_t)$, increasing and concave.
- No discounting: $U = u_1 + u_2$.
- Suppose that $\gamma < 1$: In period 1, consumers care equally about
  - period-1 news about period-1 consumption utility
  - period-2 news about period-2 consumption utility
- But they care less about
  - period-1 news about period-2 consumption utility

Overconsumption.

- Suppose first that W is deterministic.
- Optimal strategy: maximize consumption utility, $c_1 = c_2 = W/2$.
- Will consumer choose this?
- If $\gamma < 1/\lambda$, optimal to deviate, so (without commitment) $c_1 > W/2$.
- Intuition (first in Stone, 2005):
  - Give self surprise gain in current consumption at cost of what is now a surprise loss in future consumption.
  - But once future rolls around, absorbed into reference point.
  - If γ low, bad news about future doesn't bother you.

Prospect-Theoryish Stuff

Behaviorally and welfare-wise similar prediction to present bias and to temptation disutility (Gul and Pesendorfer 2001).

But:

- No "our theories same, therefore mine is better."
- No "only possible one thing can cause similar behavior."
- No horse race.

And:

- Yes to proposing identification, then measure all.
- One major prediction of $\gamma < 1$ not in PB or TU:

Overconsumption exacerbated by ex ante uncertainty.

- When uncertainty resolved right before initial consumption, then:
- Even though uncertainty resolved before first choice, can increase consumption above level for **all** wealth levels if known in advance.

Prospect-Theoryish Stuff

Intuition?

- No uncertainty, deviating to consume too costly.
- With uncertainty , good realization W, deviating partly (merely) foregone again future consumption.

Or:

- May be able get self to jog in good weather if weather always good.
- But if chance tornados, then partly loss to jog, and loss of future health only partly loss.
- Harder to follow through.

But another lesson:

- May no longer want commit!
- Some of it consistency problem
- But good behavior more unpleasant when sometimes not required ...
- If we mistake people for naive present bias, may force commitments that make them worse off.

Prospect-Theoryish Stuff

Response to unexpected wealth shocks:

- Limiting case of low-probability events:
  - suppose consumer had been expecting to have lifetime wealth W with probability one, but at the beginning of period 1 learns (to her surprise) that she actually has deterministic lifetime wealth W'.
- Independently of γ, she will respond asymmetrically to these news:
  - For small windfall income, person consumes all windfall immediately.
  - But for small bad news, defers consumption loss entirely to future.
- Intuition: surprises regarding current consumption are felt more heavily than prospective surprises regarding future consumption.
- Reflects old Thaler intuition?
- But see Pagel (2013) for countervailing forces.

Prospect-Theoryish Stuff

Precautionary Savings:$\rightleftharpoons$

- Suppose uncertainty in lifetime wealth resolved only late in life.$\rightleftharpoons$
  - Consumers hate this uncertainty.$\rightleftharpoons$
  - Can't do anything about it, per se.$\rightleftharpoons$
  - But recall: gain-loss utility is proportional to consumption utility.$\rightleftharpoons$
  - The lower the marginal utility of wealth, the less painful are losses.$\rightleftharpoons$
  - Gives clear, first-order motive for precautionary savings.$\rightleftharpoons$
  - First-order effect based on 2^{nd} derivative, in contrast to 2^{nd}-order effect based on 3^{rd} derivative!$\rightleftharpoons$
  - People save to avoid painfulness of bad news.$\rightleftharpoons$
  - Major benefit being rich is neutralizing news utility over money.

Prospect-Theoryish Stuff

Summarizing Savings and Consumption in KR (2009):

- 1st-order precautionary savings based on 2nd derivative
  - vs. 2nd-order precautionary savings based on 3rd derivative.
- Overconsumption due to time-inconsistencies
  - exacerbated by ex ante uncertainty.
- Consume unexpected wealth shocks, delay unexpected income drop.

Prospect-Theoryish Stuff

Some more economic implications:

- Taxi cabs, bicycle messengers, baseball vendors and labour elasticity. etc. Camerer et al (1997) and others:
  - Quit early when get unexpectedly high earnings (met target)
  - But work harder (and show up) on days *expect* high wages
- Mas (2006), "Pay, Reference Points, and Police Performance":
  - worker morale effected by the *surprise* element in compensation—as identified by final-offer arbitration rulings
- Card and Dahl:
  - more domestic violence on days sports team *unexpectedly* loses.
- Principal-agent models.
  - Herweg, Muller, and Weinschenk (AER, 2010): incentive schemes won't use all the information about effort to fine tune compensation.

Prospect-Theoryish Stuff

For those of us (2 people? 1?) who have drunk Koolaid on news utility, see confounds everywhere...$_{\rightleftharpoons}$

- Social comparison$_{\rightleftharpoons}$
- Habit formation$_{\rightleftharpoons}$
- Anticipatory utility$_{\rightleftharpoons}$
- Present bias, temptation$_{\rightleftharpoons}$

Grandiosity...but disciplined grandiosity.

$\looparrowright$

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.

$\hookrightarrow$

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.

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.

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. Why?

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

Consensus empirical fact: people often cooperate with others in one-shot prisoner's dilemma, but *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.

- 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.

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?

$\looparrowright$

Social Preferences

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

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

Note:

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

These are of course unrealistically extreme forms.

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.

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

Exaggerates the kinkiness of preferences.

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

Social Preferences

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

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

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"

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.

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?

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

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.)

Significant sacrifice to avoid coming out behind? $\rightleftharpoons$

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	*(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 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.

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 $\bar{\sigma} > 0$, 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.

$\looparrowright$

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?

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

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%*

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.

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.**

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.

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 better 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.

$\looparrowright$

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.

$\looparrowright$

- 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?$\rightleftharpoons$

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

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

Social Preferences

- 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). (So 10% are either silly or nasty)
- 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.

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.

Final point ....

- "Wiggle room → non-selfish behavior may be less important than it appears" is not an awful intuition.
- But it's not as compelling as may seem.
- 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.
- Bargaining selfish people vs. bargaining altruists vs. bargaining self-righteous people?

Modeling Reciprocity? Hard.

- 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.
- 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.
- How might we capture the role of volition and intentions in models of social preferences?

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.

$\looparrowright$

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?