

Social Preference Experiments: Dictator, Ultimatum & Trust Games

Joseph Tao-yi Wang

1/2/2013

Fairness, Reciprocity, and Trust

- Example: Ultimatum Game
 - Proposer: makes a take-it-or-leave-it offer
 - Responder: accepts or rejects the offer
- Why should we care about this game?
- People talk about Fairness and Trust
- Stigler (1981) “self-interest theory will win.”
- Results = price tag on negative reciprocity

Political History Example

- Federal Convention 1787, Philadelphia
- “Should new states be 2nd rate states?”
 - George Mason: “They will have the same pride and other passions which we have, and will either not unite with or will speedily revolt from the Union, if they are not in all respects placed on equal footing with their brethren...”
- Fear of rejection or Fairness?
 - Can we apply this to China-Taiwan relations?

Self-Interest or Not?

- Self-interest: What we first learn in principles
- What about altruism?
- Standard response:
 - Monetary payoff of your “friends” enter into your utility function (so you still Max. U)
- Why don't we see this later?
 - Because the 1st Welfare Theorem will fail!
- Do people really only care about their own payoffs?

Games on Social Preferences

- Prisoners' Dilemma (PD, 囚犯的兩難)
- Public Goods Game (PG, 自願捐輸賽局)
- Ultimatum Game (最後通牒談判)
- Dictator Game (獨裁分配): responder cannot reject
- Trust Game (互信賽局): Dictator game where responder invests first to determine pie size
 - Measure of **Trust**: Amount of investment
 - Measure of **Trustworthiness**: Amount of repayment
- Centipede (蜈蚣賽局): Multi-stage trust game
- Gift Exchange (禮尚往來): Multiplayer trust game

Prisoners' Dilemma (PD)

	C	D
C	H, H	S, T
D	T, S	L, L

$(T > H > L > S)$

- Each player pick C or D
- The Dilemma:
 - Both cooperate (C) is Pareto dominant
 - But, defecting (D) against C is better
- Only Equilibrium: (D, D)

Prisoners' Dilemma (PD)

- 1-shot games Baseline:
 - Play C 50% of the time
- Changing payoffs:
 - Lowering T (raising S) increases cooperation
- **Pre-play communication** raises cooperation
- Random Re-Matching:
 - Dwindle to only few cooperate

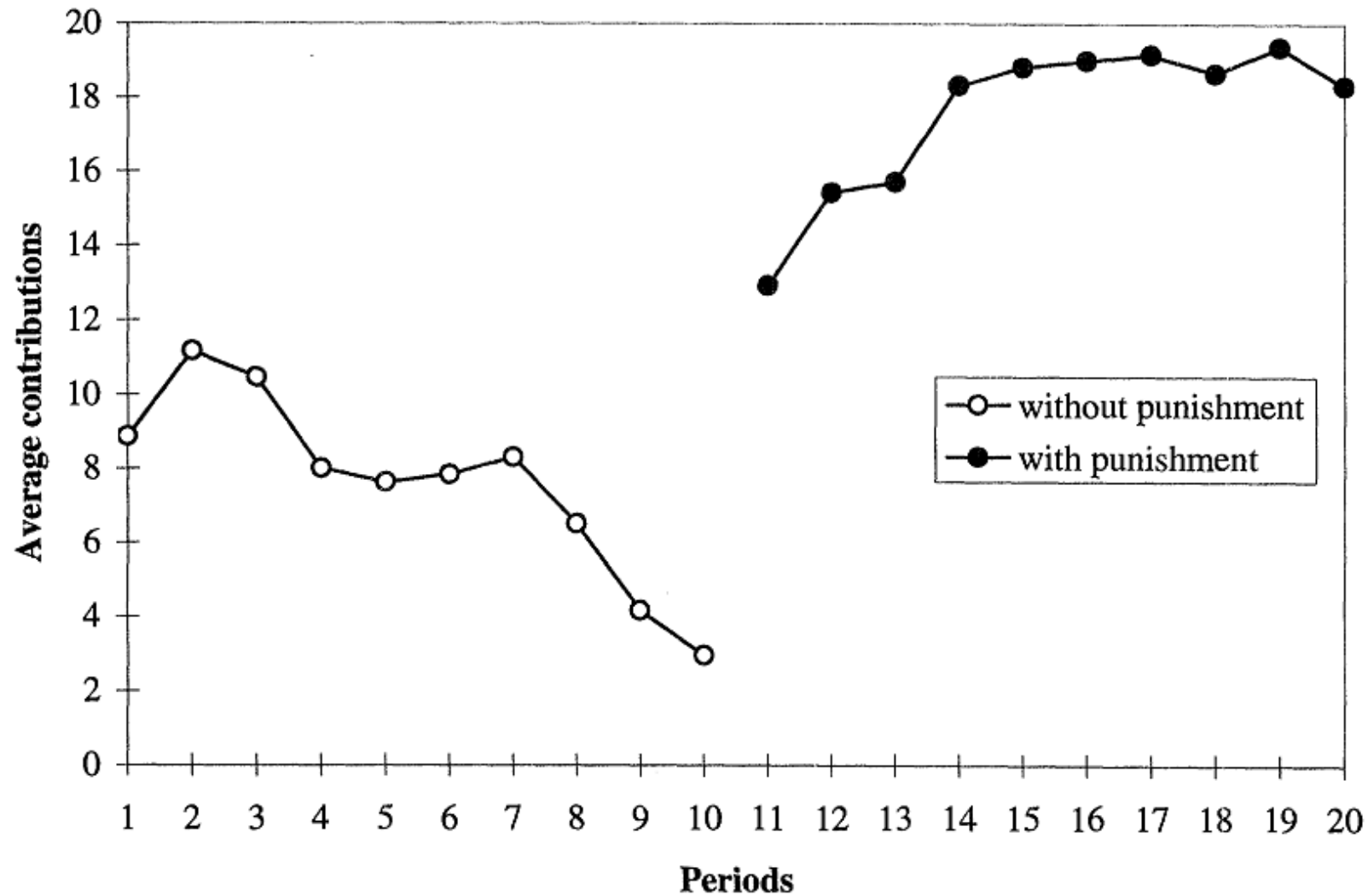
Public Goods Game (PG)

- N players
- Invest c_i from personal endowment e_i
- Total contribution $c_{all} = \text{sum of } c_i$
- Payoff = $e_i - c_i + m * c_{all} / N$
 - Total contribution is multiplied by m and divided among all players
- Like PD:
 - Cooperation is good; want to free-ride

Public Goods Game (PG)

- 1-shot games Baseline:
 - Average contribution = 50% (mostly all or none)
- Changing payoffs:
 - Raising m (marginal return) raises contribution
- **Pre-play communication** raises cooperation
- (Random) Re-Matching: Contribution dwindles
- Punishment Effect: (Fehr and Grachter, AER 2000)
 - Even though one can free ride other's punishing
- Cooperation seems to be **Reciprocal**

Fehr and Grachter (AER 2000)



Pure/Impure Altruism

- Example: $U_i(X) = x_i + \alpha \cdot x_{-i}$
- Can explain (C,C) in PD
- Homework: Can this explain PG (with or without punishment)?
- Altruistic giving is crowded out if others give
- Can't explain **reciprocity**
 - “I like to do good to those good to me, but do bad to those bad to me.”

Inequality-Aversion: Guilty-Envy

- Fehr and Schmidt (1999)

$$U_i(X) = x_i - \frac{\alpha}{n-1} \sum_{k \neq i} \max(x_k - x_i, 0) \\ - \frac{\beta}{n-1} \sum_{k \neq i} \max(x_i - x_k, 0)$$

- Envy stronger than guilt: $0 \leq \beta \leq 1, \beta \leq \alpha$
- Explains
- PD: sustain cooperative outcome (C,C)
- PG: heterogeneous contribution
 - some 0 vs. some positive
- Punishments in PG

Classroom Public Goods Game

	回合數1			回合數2		
集合總張數	13			5		
集合分配得分	2.6			1		
留下張數	0	1	2	0	1	2
累積組數	3	5	5	0	1	12
本回合分數	2.6	3.6	4.6	5.6	6.6	7.6
累積總數	3	5	5	3	6	4

PD and PG: Conclusion

- Experts in these two games:
 - PD: Chun-Lei Yang (Academia Sinica)
 - PG: Li-Chen Hsu (NCCU)
- Do these results falsify game theory?
 - Not quite. They invite for new theory
- New theory in BGT: **Social Preferences** (BGT, 2.8) and **Limited Strategic Thinking** (BGT, Ch.5)
- Problem with PD/PG: “Defecting” is dominant
 - Can’t distinguish altruism from conditional cooperation

Ultimatum Game

- A “Better” Game: Ultimatum Game
 - Proposer: makes a take-it-or-leave-it offer
 - Responder: accepts or rejects the offer
- **Baseline:** 1-shot, anonymous, action
 - Random re-matching
 - Strategy Method: Minimum Acceptable Offer (MAO)
- Strategy Method vs. Specific-action Method
 - Is the strategy method too “unnatural”?

Ultimatum Game

- Basic Results (BGT, Table 2.2, 2.3)
- Proposer
 - Mode / median: 40~50%
 - Mean: 30-40%
 - Almost no below 10% or above 50%
 - **Fairness** or **Fear of Rejection**?
- Responder
 - Rarely reject offers of 40~50%
 - 50% rejection rate for offers below 20%

Inequality-Aversion: Guilty-Envy

- Fehr and Schmidt (1999)

$$U_i(X) = x_i - \frac{\alpha}{n-1} \sum_{k \neq i} \max(x_k - x_i, 0) \\ - \frac{\beta}{n-1} \sum_{k \neq i} \max(x_i - x_k, 0)$$

- Envy stronger than guilt: $0 \leq \beta \leq 1, \beta \leq \alpha$
- Explains
- Ultimatum – Rejections, Fair offers
- Can Altruism also explain rejection / offers?
 - No (so it's less “parsimonious”; inferior to G-E!)

ERC (Envy, Reciprocity, Competition)

- Bolton and Ockenfels (2000)

$$U_i(X) = U \left(x_i, \frac{x_i}{\sum_{k=1}^n x_k} \right)$$

- Care about **relative** share
- No **individual comparison**; Only **total comparison**
- Homework: Can this model also explain PD and Ultimatum rejection/offers?
- What game can distinguish this from G-E?

Dictator Game

- An Ultimatum Game with rejection
 - Proposer: makes a dictated allocation decision (and the Responder cannot reject it)
 - Distinguish **Fairness** from **Fear of Rejection**
- Basic Results: (BGT, Table 2.4)
 - Lower than Ultimatum, but not zero
- Offers are more generous than BR
 - Both **Altruism** AND **Strategic Concerns** exist
 - Proposers hold “pessimistic” belief

Dictator Game: Guilt-Envy

- Guilt-Envy Prediction for the dictator game:
 - Give 50-50, or nothing
- Not consistent with the dictator game results
 - Homework: Try to “fix” this by adding concavity
- What are more plausible theories?
 - ERC: More sophisticated theory of altruism
 - Fear of rejection + Self-interest
- Homework: Show how ERC or “fear of rejection” can explain dictator game results

Can XYZ... explain these results?

- X: Methodological Variables
 - Repetition, **Stakes**, Anonymity & Experimenter “**Blindness**”
- Y: Demographic Variables
 - Gender, Race, Academic Major, **Age**,
 - Brains, Biology and **Beauty**
- Z: **Culture**
- XX: Descriptive Variables
 - Labeling and Context
- YY: Structural Variables
 - **Add a move** (see below)

X: Methodological Variables

- Repetition
 - Experience effect (low offers/rejects) is small
 - Unless played with self-interest robots
 - Is the small effect a “satiation” of emotion? Try “restart”
- Stakes
 - Very large changes (N month wage) only have a modest effect on rejections; no effect on offers
 - Match contribution: $\frac{1}{2}$ selfish, $\frac{1}{3}$ Leontief, $\frac{1}{6}$ utilitarian
- Anonymity and **Experimenter “Blindness”**
 - Demand effect vs. Double-blind design:
 - Mean is 10%; half gave 0 in dictator; no effect in ultimatum

Y: Demographic Variables

- Gender – No simple “main effect”
 - Reject less in ultimatum; “punish smartly”
- Race – Few results (political correctness?)
 - Ex: White male repay less to Asians! (Is this social status?)
- Academic Major – Mixed results on Econ-Majors
- **Age** – Self-interest → strict equality → equity
 - EX: Kindergartners accept 1 penny 70% (vs. 30-60%)
- Brains, Biology and **Beauty**
 - Strongness: High-T reject more, but offer generously
 - Handsomeness: Many women give >50% to attractive men

Z: Culture

- Methodological Issues (and Solutions):
- **Stakes**: Equal purchasing power; N-day wage
- **Language**: “Back translation”
- **Experimenter Effects**: Bilingual, seen as equal
 - Each should run a session in one culture
- **Confounds**: Match two cultural samples on demographics and measure “uncontrollables”

Z: Culture

- Machiguenga Farmers in Peru
 - Offer average 26%; mode 15%
 - Social disconnect; no names for non-relatives
- Henrich et al. (2002): 20 cultural groups
- Hyperfair offers (that are rejected!)
 - Ache headhunters of Paraguay and Lamelara whalers of Indonesia
 - Competitive gift-giving: Accepting a hyperfair offer incurs obligation to repay and is an insult

Z: Culture

- Two key determinants ($R^2 = 0.68$):
 - Amount of **cooperative activity** (economies of scale in production)
 - Degree of **market integration**
- More cooperative activity and market integration lead to 50-50 sharing norms
 - Active markets and self-interest don't sync!
- This is a real culture study...

XX: Descriptive Variables Labeling & Context

- **Self-interest is okay in the market:** A buyer-seller story lowers ultimatum offers by 10% but does not affect rejection rates
- **Claiming shared resources creates common ownership:** Both become more generous
- **Priming:** Prompting instructions (“What would you do if you were the other side?”) increase fear of rejection
- General Principles of Framing? TBD!

YY: Structural Variables

- **Identity, Communication, Entitlement**
 - Dictators more generous if identity of recipient is known (student, well-known charity,...) or introduced
 - Winning the right to propose lowers offers by 10%
- **Competitive Pressure and Outside Option**
 - Offer less if “need to earn X to go to next round “
 - $U(\text{reject}) = (2,3)$: Multiple focal points; disagreement
- **Information about the Amount being Divided**
 - Incline to reject since low offer “could” be fair
- **Multiperson Games**: Competition drive offers
- **Intentions**: What if only two options $(8,2)$ & $(10,0)$?

Can G-E, ERC, Explain These?

- **Homework:** Show how G-E can explain ultimatum game results with competing proposers or respondents
- **Homework:** Show how ERC predicts “ignoring allocation to the inactive Recipient” in Guth and Van Damme (1998):
 - Proposer offers (x, y, z) to Two Responders
 - “Active” Responder sees y and/or z ; accepts or rejects
 - See y or yz : Offer $y=30-40\%$, $z=5-10\%$, rejection $\sim 5\%$
 - See only z : Offer $z=12-15\%$, keep most, rejection $\sim 5\%$

ERC vs. Guilt-Envy

- Other games: ERC don't match data as well as G-E
 - People care about inequality among others: Charness and Rabin (2000), punishment in PG
 - Absolute difference still matters: (see below)
- But, both models assume **separability**
 - Are utilities of terminal-node payoffs separable from game tree path and unchosen payoffs? (Some evidence against this, but might be a good approximation)
- Both models don't capture **reciprocity**
 - “I like to do good to those good to me, but do bad to those bad to me .”

Fairness Equilibrium

- Psychological Games: Rabin (1993)
- Normal Form Games; Action: a_1
- Belief about other's action: b_2
- Belief about belief: c_1
- 1 's kindness toward 2 :

$$f_1(a_1, b_2) = \frac{\pi_2(b_2, a_1) - \pi_2^{\text{fair}}(b_2)}{\pi_2^{\text{max}}(b_2) - \pi_2^{\text{min}}(b_2)}$$

Fairness Equilibrium

- 1 's kindness toward 2 :

$$f_1(a_1, b_2) = \frac{\pi_2(b_2, a_1) - \pi_2^{\text{fair}}(b_2)}{\pi_2^{\text{max}}(b_2) - \pi_2^{\text{min}}(b_2)}$$

- 1 's perceived kindness of 2 :

$$\tilde{f}_2(b_2, c_1) = \frac{\pi_1(c_1, b_2) - \pi_1^{\text{fair}}(c_1)}{\pi_1^{\text{max}}(c_1) - \pi_1^{\text{min}}(c_1)}$$

Fairness Equilibrium

- Player 1's (social) preferences:

$$U_1(a_1, b_2, c_1) = \pi_1(a_1, b_2) + \alpha \cdot \tilde{f}_2(b_2, c_1) \\ + \alpha \cdot \tilde{f}_2(b_2, c_1) \cdot f_1(a_1, b_2)$$

- Rational expectations: $a_1 = b_1 = c_1$
- Example 1: PD
- Example 2: Chicken Game
- Extensive-Form Fairness Equilibrium
 - Falk and Fischbacher (1998)

ERC, Guilty-Envy vs. Fairness Equilibrium

Offer	Accept	Reject	Reject (%)	ERC	G-E	Fairness Eq.
Equal	5,5	0.5, 0.5				
Unequal	8,2	0.8, 0.2	38	None	Some	Some
Equal	5,5	3,3				
Unequal	8,2	6,0	19	None	None	Some

What did we learn from all this?

- A LOT has been done...
 - Is there a parsimonious theory to explain all?
- Every stone has been turned to disprove Social Preference, but “failed”...
 - People are not strictly self-interest
- Methods: See how careful they did those!
- What makes a result interesting?
 - How can you adopt it in your own design?

Trust

- How do I know you will hold up your end of the deal?
 - Legal Contracts, Third-party assurance,
 - Family solidarity, threats of violence
- These are costly; **Trust** is cheap!
 - Lending a truck to strangers in Iowa
 - Tokyo's lost and found center (72% returned)
 - Firms prefer to lay off rather than cut wages

Trust

- **Trust** (Social Capital) explains growth
- Putnam (1995)
 - “Since trust is so central to a theory of social capital, it would be desirable to have strong behavioral indicators of trends in social trust or misanthropy. **I have discovered no such behavioral measures.**”

Trust Game

- Investor decides how much to invest
 - Amount invested is multiplied by m
- Trustee decides how much repay investor
- How much would you invest? How much would you repay?
- This game provides a measure of Social Capital:
 - Trust: Amount invested
 - Trustworthiness: Amount returned

Trust Game

- Berg, Dickhaut and McCabe (1995)
- Double-blind; \$10, $m=2$
- Investor put in about 50% of endowment
- Trustee repay about 95% of what was invested
- Replications: Various papers
 - Invest 40-70%
 - Repay 110-150% (Lowest: “corrupted” Kenya - 55%)
- Exception: Van Huyck, Battalio Walters (95/01)
 - Peasants (invest little) vs. Dictator Landlords (take all)

Trust Game: Why Trust?

- Is Trustworthiness **Reciprocity** or **Altruism**?
- **Altruism**: Dictator game offer
- **Reciprocity**: Difference between “% repay in Trust game” and “% Dictator game offer”
- Dufwenberg and Gneezy (2000):
 - 30% vs. 33% (insignificant) - **Altruism** alone?
- Cox (1999):
 - 10% more (statistically significant but small)

Trust Game: Direct vs. Indirect

- Matching Design:
 - Pair → Foursome → Society
- Buchan, Croson and Dawes (2000)
 - Trust: 64% → 48% → 39%
 - Trustworthiness: 35% → 19% → 20%
- Dufwenberg et al. (2000)
 - Trust: 60% → 53%
 - Trustworthiness: 28% → 37%

Trust Game: Other Extensions

- History Effect in sequential trust game:
 - Donate 250 (at cost 150)
 - See past 6 rounds donation history
- Seinen and Schram (1999)
 - 25% \rightarrow 70% (Show Donor History)

Trust Game: Other Extensions

- Multistage Trust Games
 - Like centipede games: but terminal node= $(0,0)$
 - **Selfish guys** can't mimic **nice guys** who all pass
- **Ho and Weigelt** (Management Sci. 2005)
 - 4 moves; pass doubles pie; strategy approach
 - 30% (50%) player 1s (2s) take 95% at first node
- **Rapoport et al.** (GEB 2003):
 - 3 person; 9 nodes; up to \$1,500
 - 1/3 of the games ended at the first two nodes

Trust at Work: Gift Exchange?

- Fehr et al.; cf. Fehr and Gächter (JEP 2000)
- 8 workers and 6 firms
- Firms offer wage w to worker (suggest e')
- Workers (if accept) chose effort e
- Payoffs: Firms earn $(q-w)e$
 - Workers earn $w-c(e)$
 - $c(\cdot)$ convex on $0.1 \sim 1.0$
- What would you choose/offer?

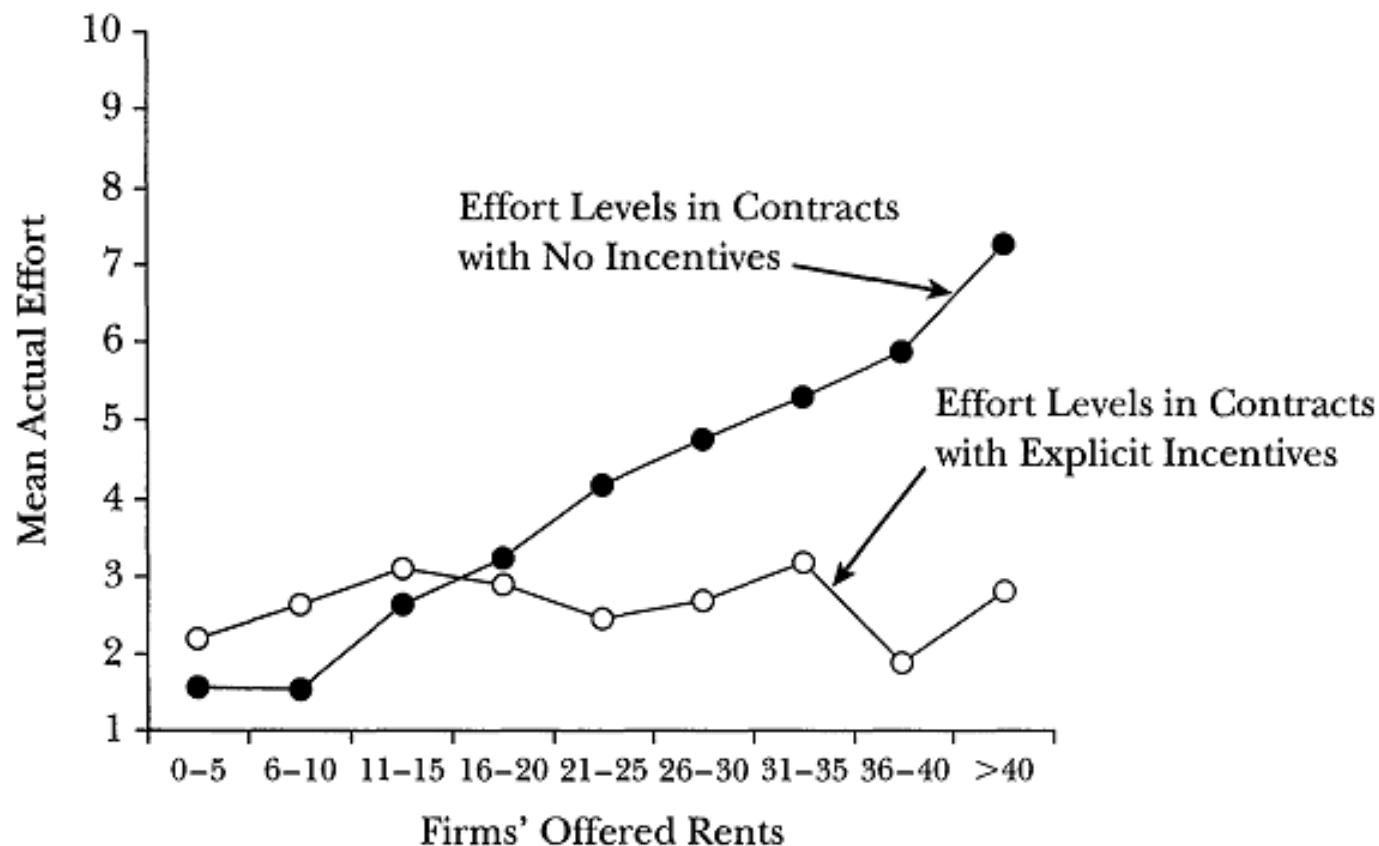
Trust at Work: Gift Exchange

- Standard Game Theory Predictions:
 - Workers will choose minimum effort
 - Firms offer low wage
- Gift Exchange (Akerlof, 1982)
 - Workers reciprocate high wage with high effort
- Efficient Wages
 - High wage creates a **job rent of $w - c(e')$** that workers might lose if they are caught shirking
- Hard to separate GE and EW in the field...

Trust at Work: Gift Exchange

• Fehr and Gächter (JEP 2000) support GE

Figure 3
Actual Effort-Rent Relation in the Absence and Presence of Explicit Performance Incentives



Trust at Work: Gift Exchange

- Other papers show less gift exchange
- Gneezy and List (2006): Initial gift exchange effect goes away after a few hours in the field
 - “Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments,” *Econometrica*, 74(5), 1365-1384.
 - “We experimented with the individual-specific variables found to be important in Landry [et al.] (2006) and found that their inclusion **does not change the qualitative insights.**”

Trust Game: Where we stand?

- How robust is gift exchange in the labor market (experimental or empirically)?
- This is still an active field of research
- Question: Where does trust come from?
- Kosfeld, Heinrichs, Zak, Fischbacher and Fehr (2005), "Oxytocin increases Trust in Humans," *Nature* 435, 2 June 2005, 673-676.

Conclusion

- Do people respond to incentives?
 - Yes! But what kind of incentives?
- External (monetary) Incentives: Payoffs
- Internal Incentives: Fairness, Altruism, etc.
- Plenty of experiments on social preferences
 - Don't blindly propose to run another one!
 - Check literature first! (BGT, ch.2, MGS, ch.12-14)
- Is there a parsimonious theory to explain all this (and make new predictions)?