

14.11 Spring 2006: Putting Social Science to the Test.  
Lecture Note #5: Fairness

David Autor

March 14, 2006

## 1 INTRODUCTION

In this lecture, we'll be continuing the discussion from lecture #3 on incentives and cooperation. In that lecture, we considered how people respond to incentives. In this lecture, we'll ask to what degree considerations of 'fairness'—or, what economists typically call 'social preferences'—shape behavior. [Why the term 'social preferences'? Because the preferences we are discussing are generally thought of as other-regarding rather than entirely self-regarding, as in the basic economic model.] First, we'll consider some influential papers presenting evidence that social preferences have significant effects on behavior, including in high-stakes transactions.

Next, we'll consider two recent papers that critique this line of work. These critiques will focus on external validity, a core methodological topic that is also at the heart of today's lecture. External validity is the extent to which the results of a study can be applied to circumstances outside the specific research setting in which the study was carried out, i.e., do the results provide useful information about 'the real world.' It is quite possible to conduct a study that is internally valid but has very limited external validity in that its results do not easily generalize to any set of circumstances outside of those created by the study. Q: Can a study have external validity without having internal validity?

## 2 FAIRNESS AND ENTITLEMENTS: KAHNEMAN, KNETCH AND THALER (1986)

Let's begin with a classic study on perceptions of fairness. The basic principal that KKT propose here is a 'dual entitlement' model of transactions in which consumers are 'entitled' to a transaction at the terms given by a 'reference' transaction and firms are entitled to a profit equal to a 'reference profit.' What precisely constitutes the reference transaction is a slippery concept—but that's in some sense the point of the article: by changing the *reference* transaction, one can cause the same *action* to be perceived quite differently (in terms of fairness). We'll see many clever examples of this in the paper. Survey questions 1 through 3 in the paper demonstrate the power of reference transactions to shape perceptions of fairness.

Q: What is the 'reference' transaction rule in economic models? Is this reference transaction subject to re-framing?

The second section of the paper incorporates the notion of 'loss aversion'—averting losses is considered more justifiable than exploiting gains (note: these are isomorphic in economic theory). Inter-

estingly, these perceptions of losses are also susceptible to framing.

This work is exceedingly clever and provocative. And the authors draw bold theoretical propositions based on their findings. But the work is certainly subject to criticism. The most important criticism concerns external validity. KKT conduct a survey of *attitudes*; they do not provide a test of behavior. It is very easy for survey respondents to condemn or condone actions as ‘fair’ or ‘unfair’ since it is costless to do so. But we have no idea of the economic ‘worth’ of these beliefs. Will consumers punish hardware stores that raise the price of shovels when it snows? Will workers quit rather than take an ‘economically justifiable’ pay cut. Will firms reduce bonus pay by 10% but not reduce base pay by the same amount? Or are all of these complaints about fairness just ‘cheap talk?’

To answer these questions, we need evidence on ‘fairness’ from a setting where there are meaningful economic stakes.

### 3 FEHR, FISCHBACHER AND TOUGAREVA (2002): “DO HIGH STAKES AND COMPETITION UNDERMINE FAIRNESS? EVIDENCE FROM RUSSIA”

[Quoting from List and Gneezy 2006] “Neoclassical economics treats labor as a hired input in much the same manner as capital. Accordingly, in equilibrium, the firm pays market-clearing wages and workers provide minimum effort. The validity of this assumption is not always supported by real life observations—some employers pay more than the market-clearing wage, and workers seemingly invest more effort than necessary (Akerlof, 1982). The “fair wage-effort” hypothesis of Akerlof (1982) and Akerlof and Yellen, (1988; 1990) extends the neoclassical model to explain higher than market clearing wages by using a gift exchange model, where “On the worker’s side, the ‘gift’ given is work in excess of the minimum work standard; and on the firm’s side the ‘gift’ given is wages in excess of what these women could receive if they left their current jobs” (Akerlof, 1982, p.544).

The gift exchange model is based on the critical assumption of a positive relationship between wages and workers’ effort levels. Workers are assumed to respond to high wage levels by increasing their effort (positive reciprocity) and to low wage levels by decreasing their effort (negative reciprocity) to the minimum required, in retaliation for the low wage. A large and influential body of empirical evidence in support of reciprocity has been reported in the past two decades. One of the first experiments to test this assumption is Fehr et al. (1993), who constructed a market with excess supply of labor, ensuring a low equilibrium wage. Under their setup, employees had no pecuniary incentive to raise

the quality of their work above the exogenously given minimum. If an employer expects employees to invest only the minimum effort required, then she has no compelling reason to pay wages above the market-clearing level. Contrary to this prediction, however, most employers in the Fehr study attempted to induce employees to invest greater effort by offering them higher (at times by more than 100%) than market-clearing wages. On average, this high wage was reciprocated by greater employee effort, making it profitable for employers to offer high wage contracts. Subsequent laboratory exercises have largely led to similar conclusions.”

The 2002 paper by Fehr, Fischbacher and Tougareva responds to the two most common criticisms of conclusions drawn about the importance of fairness and ‘social preferences’ to economic behavior. Those criticisms are: (1) the stakes are too small, so it is near-costless to indulge ‘soft-headed’ norms like fairness; (2) the experimental settings often lack a competitive market environment, which may tend to discipline or displace indulgence of ‘social preferences.’ This paper introduces both high stakes and competition into an environment where fairness has scope to operate.

**The experiment takes place in a ‘Gift Exchange Market’ (GEM) with the following structure:**

- In each round, there are 9 workers and 6 firms. ‘Firms’ and ‘workers’ are experimental subjects randomly assigned to these roles.
- Interactions are completely anonymous. Firms and workers cannot develop reputations.
- Each firm can hire up to one worker. Hence, there is always *excess supply* of labor (9 workers vs. 6 firms). Firms have no direct economic incentive to pay workers more than their ‘leisure’ wage since there will always be involuntary unemployment in this market (assuming every subject wants to work).
- The structure of the market is a one-sided oral auction in which firms post wage offers and workers choose whether or not to accept them.
- A firm cannot make an offer to a specific worker—any worker can accept. Once the offer is accepted, a labor contract is concluded (there is no renegotiation).
- A firm can continue to improve its offer until accepted.

- Wage offers do not stipulate effort levels, but workers do have a choice of ‘effort,’ explained below.
- Once workers choose effort, worker and firm payoffs are calculated and a new round begins (that is, there is no real work effort here).
- The firm’s payoff from a concluded wage contract is:

$$\Pi = (120 - w) e,$$

where 120 is a constant chosen by the experimenter and known to the firm,  $w$  is the offered wage and  $e$  is the worker’s chosen effort level.

- The worker’s payoff from a concluded wage contract is:

$$Y = w - c(e) - 20,$$

where  $w$  is the wage,  $c(e)$  is the worker’s cost of effort (given next) and 20 is a fixed price of working (needed to give experiment control of size of payoff).

- The effort cost function is shown in the accompanying slides. The monetary cost of effort is an increasing, convex function such that each marginal unit of effort costs more than the last for the same increment to firm profits.

In addition to the GEM, there is also a Complete Contracts Market (CCM), where effort is fixed exogenously at the lowest level. There is no scope for reciprocity in the CCM.

What does competitive theory predict will occur in the GEM? At any given wage offer, an income-maximizing ‘worker’ will choose effort of 0.1 at cost 0. Any other choice would reduce  $Y$ . Recognizing this, firms should offer the minimum wage needed to bring workers into the market, which is 20.

**Stakes:**

- The experiment was conducted in Austria with standard low stakes (something like \$10 average earnings per participant).
- The experiment was conducted in Russia with low stakes (a purchasing-power equivalent sum).

- The experiment was conducted in Russia with high stakes. On average, subjects earned \$40 to \$50 U.S. dollars in a high stakes GEM. The median monthly income of their Russian subjects was \$17. Thus, this is two to three times monthly income.

We would be surprised if the low-stakes experiments did not generate behavior at odds with the competitive predictions, since ‘fairness’ has been observed in many laboratory settings. The question is whether such fairness survives the high stakes setting, where workers may be giving away a week or two of pay by reciprocating and firms may be giving up a similar amount by making wage offers above the minimum level (which is 20).

**Results:**

- Figure 1 demonstrates that the CCM behaves much like the theory predicts. The average wage starts at about 35 but drifts downward towards 20 and, in the final period, averages only 27.
- In the GEM with both low and high stakes in Russia, the agreed wage starts at close to 55 and rises towards 60 to 65 in the final periods.
- Figure 2 shows that effort is strongly increasing in the agreed wage, consistent with gift exchange. The slope is a little shallower in the high stakes condition, but there is no question that effort is responding to the wage in both high and low stakes settings.
- Figure 3 shows that in the normal stakes condition, Russians and Austrians obtain similar wage levels in the GEM. This is important as a check of external validity (i.e., checking that Russians are not unusually cooperative). The high stakes condition is not used in Austria because of the expense.
- Figure 4 shows that Russian and Austrian subjects have comparable effort responses to wage levels.

The conclusion of this study is therefore very straightforward: high stakes and market competition do not drive out cooperation (or ‘social preferences’) in the marketplace.

#### 4 METHODOLOGICAL INTERLUDE: EXTERNAL VALIDITY

In a previous lecture, we have discussed some of the major threats to *internal* validity. A quick review, these are:

- Non-compliance
- Attrition
- Externalities
- Contamination of control group (of which externalities are one example)

Now, let's consider some threats to external validity. There are a number of classic confounds that cause an experiment to produce results that, while internally valid, are not generalizable to the external world:

- Hawthorne effect—When the mere fact of observation or attention causes the treatment group to change its behavior (could be called a ‘Heisenberg’ effect). Named after a series of studies designed to raise productivity at Western Electric in 1939 by Hawthorne et al. where (it was believed) that treatment group subjects increased their productivity not because the intervention was intrinsically useful but because the attention they received caused them to work harder. Note that the study still has *internal* validity in that treatment *caused* a productivity increase. But the external applicability of this finding may be nil.
- John Henry effect—Where public assignment of units to treatment and control groups may cause control group members to engage in social competition to show it can perform as well as the treated group despite not receiving treatment. This term is named after John Henry, a steel driver who, when he knew his output was to be compared with that of a steam drill, worked so hard that he outperformed the drill (and died).
- Demand effects—In which subjects or respondents “follow orders” or cooperate in ways that they almost never would under their routine daily lives. An extreme case is in the infamous Stanley Milgram experiments in which treatment subjects administered what they thought were potentially lethal electric shocks to their peers at the orders of the experimenter.

- More generally, there are a variety of Social Desirability effects:
  - Subjects may become nervous about being monitored, or evaluation apprehension. When people become anxious, many things happen. Physiological indicators, such as heart rate or blood pressure, change. If people are slightly anxious, they may do better on tests, performance, or assessments. However if people are very anxious ("flooded") they will almost certainly do worse.
  - People may attempt to appear smarter, more attractive, or more tolerant than they normally are. Paper and pencil questionnaires are especially prone to these effects because often the answers are not checked for their veracity. (And, on online surveys, we may not even correctly know who anyone is.)
  - It is not just individuals who have social desirability effects. A century ago, the Russian writer Lev Tolstoy described "Potemkin villages." When the Czar went on cross-country trips, government officials were a little ways ahead of him. In cooperation with local government, they would erect false-fronted buildings (as on a movie set) and the best looking young men and women of the village would stand before these fake structures smiling and throwing flowers. While most groups or organizations will not go to this extent, they may "hide" their more embarrassing members, "fudge" or slightly alter records to appear more virtuous.

Other core issues:

- Population – Is the population in which this was carried out particular, or would this result extend to other programs?
- Strange populations – Many experiments have been performed with baseball card traders. Who are these people?
- Robustness – Was the program run better because it was an experiment than it would be under normal circumstances?
- Time span – How does the long run effect (generally not observed) compare to short run (typically measured)?

- General Equilibrium – If the program was implemented on a large scale would there be GE effects? Would these augment or attenuate the individual level effects?
- Definition of treatment – Was the program a combination of features which do not allow us to say whether it would work again unless all features were replicated?
- General artificiality of the laboratory. How much do we learn from lab experiments generally?
- Irrelevant choices – Baseball cards, mugs. Do these choices correspond to anything in life?

## 5 EXTERNAL VALIDITY IN EXPERIMENTS: LAZEAR, MALMENDIER AND WEBER (2006)

You might think that the Fehr et al. paper puts concerns about external validity to rest. After all, these authors have tested fairness in a labor market setting (of sorts) and have tested it in a setting where the stakes are extraordinarily high. The fairness hypothesis appears robustly confirmed by these tests. In this paper, LMW offer an insightful critique. Their insight is that experiments do not simply create an artificial environment (e.g., a labor market without any labor) but also generate an artificial set of participants.

A basic insight of economics is that markets induce self-selection; they pair lowest cost sellers with highest value buyers who transact with one another. Buyers and sellers of goods are emphatically *not* individuals placed in a market at random; they self-select into the marketplace because they expect to benefit from these transactions. Those who would expect not to benefit will generally stay home. Compare this to a typical laboratory experiment (like the one above), where buyers and sellers are assigned roles at random. Perhaps the participants in these experiments are *unlike* participants in a market precisely because they are selected at random. Or, to quote from their paper, an experiment with randomly selected individuals might reveal that a significant portion of subjects suffer from acrophobia. But voluntary sorting ensures that those who build skyscrapers are unlikely to suffer from this fear.

The objective of the LMW paper is to assess whether sorting is a significant explanation for the robust evidence for ‘social preferences’ found in laboratory experiments. The setting of their experiment is a standard ‘dictator game,’ in which experimental subjects are given \$10 and must decide how to allocate it between themselves and another experimental subject. The dictator has

no interaction with the recipient and the recipient has no opportunity to reject whatever division is offered. Standard theory says that dictators will give essentially nothing to recipients. But as has been widely found, the average dictator will give about 40 percent of the money to the recipient.

The twist in this study is to allow subjects to self-select between the dictator role and a purely passive ‘receive a payment’ role. In the passive role, subjects do not have to dictate and simply receive \$10. From the point of view of a purely self-interested agent, these roles are identical. But for other types of agents they are not. LMW first run the dictator game with each subject. They next run the experiment again, allowing participants to self-select between dictator and passive recipient. They finally re-run the experiment, raising the stakes in the dictator game above \$10 while keeping the stakes in the passive game at \$10.

Their conceptual framework distinguishes between three types of agents: those who like to share; those who do not like to share; and those *who do not like not-sharing*. The first category includes agents with ‘social preferences.’ The second contains neoclassical economic actors. The third category is most interesting. These are subjects who will share when placed in a setting where sharing is feasible. But given the choice, they will avoid such a setting (even potentially at a positive cost) so they do not feel obligated to share. Implicitly, these agents do not like sharing, but they do not like being seen not to share when sharing is feasible. LMW liken this difference between like-to-share and not-like-to-*not*-share agents as the difference between shame and guilt. Shame operates only when others observe the action. Guilt occurs whether or not the action is observed.

## 5.1 CONCEPTUAL MODEL

The article contains a simple and illuminating conceptual model that is worth exploring. In this model, there are three types of agents, those who: 1) dislike sharing; 2) like sharing; 3) dislike not sharing.

Consider an agent endowed with an amount  $w$  that she must allocate between herself ( $x$ ) and another agent ( $y$ ), with

$$w = x + y.$$

Let the agent’s utility depend on the allocation *and* the sharing environment:

$$U = U(D, x, y),$$

where  $D$  is a dummy variable equal to 1 if the environment allows sharing and equal to 0 if the allocation of  $w$  is exogenously determined.

The agent's propensity to share in a sharing environment ( $D = 1$ ) is given by  $\alpha$ , with

$$\begin{aligned}x &= \alpha w, \\y &= (1 - \alpha) w.\end{aligned}$$

The case of  $\alpha = 1$  corresponds to 'standard' preferences. Agents with  $\alpha < 1$  either like sharing *or* dislike *not* sharing. By implication, those who like sharing would pay to be in a sharing environment. Those who dislike not sharing would pay *not* to be in a sharing environment.

Let  $w'$  be the amount of income that an individual needs to be indifferent between the sharing condition with wealth  $w'$  and the non-sharing condition with wealth  $w$ . Define

$$\lambda(w) = w'/w.$$

For agents with  $\lambda = 1$ , preferences are standard. Agents with  $\lambda < 1$  like to share—that is, they need less money in the sharing condition to be as happy as in the non-sharing condition (note that  $w'$  is the amount they are given to allocate;  $x'$  will presumably be less than  $w'$ ). Agents with  $\lambda > 1$  dislike not sharing. They will pay to avoid the sharing environment.

Consider the following proposition: *The average amount shared is (weakly) smaller when individuals can sort between environments than when they cannot.*

Why would this be true? Those who are fully selfish have no incentive to sort. Those who like to share also have no incentive to sort. But those who dislike not sharing have an incentive to avoid the sharing environment. Since these agents share more than the fully selfish agents, their self-selection *out* of the sharing environment will reduce average sharing in that environment.

A concrete example is helpful here. You encounter a beggar on the street who asks for a handout. If you are fully selfish, you do not share and do not mind. If you like sharing, you share and enjoy it. If you dislike not sharing, you cross the street before encountering the beggar to avoid being asked to share. On the other hand, if you were confronted by the beggar, you would share. This example points out the potential external validity problem with 'sharing experiments'—we may be forcing some agents to share by confronting them with a beggar when, in real life, these agents would have crossed the street to avoid him.

In the appendix of the paper LMW develop a parametric utility function that places additional structure on their predictions. In this model, individuals share a constant fraction of their income  $\alpha_i$  when placed in the sharing environment. But among those with  $\lambda_i > 1$  (i.e., dislike not sharing), those with the lowest  $\lambda_i$  are those with the lowest  $\alpha_i$ . That is, the stingiest among those who dislike not sharing are those who require the smallest amount of compensation to be placed in the sharing environment. This has an intuitive logic: if you dislike sharing and only share a little bit when placed in the sharing environment, then probably your psychic cost of this environment is low (since, after all, you aren't giving away much money). [But the result is not entirely general—only some utility functions will have this property.]

Under this parametric form, the model has another implication: *The endowment  $w'$  at which individuals who dislike not sharing first enter the market is decreasing in  $\alpha$ .* In plain language, this means that individuals who dislike not sharing who give the *most* will also be those most likely to *avoid* the sharing environment. Return to the beggar example. Assume that avoiding the beggar now means waiting five minutes at a crosswalk to get to the other side of the street. Who will be willing to pay this price? Not selfish agents and not those who like sharing. Among those who dislike not sharing, it is plausible that those who share the most will be the most motivated to cross the street (even at a positive price) because they have the most to lose by confronting the beggar.

## 5.2 EXPERIMENTAL DESIGN

The first decision round is a (mandatory) dictator game with an allotment of \$10 in quarters that the dictator had to allocate between herself and the recipient. This is done two ways. In the no-anonymity condition, the dictator is informed that the recipient will be told after the fact which dictator made her allocation (though neither would know in advance with whom they were paired). In the anonymous condition, this information is not revealed. The reason for using the no-anonymity condition (which is non-standard) is because it appears to correspond more closely to what typically occurs outside of the laboratory.

In the second decision round, dictators are allowed to opt out of being a dictator. If they opt out, they simply receive \$10. The potential recipient is not informed of any decision—hence, this is an unobservable action.

Rounds 3 through 6 were roughly identical to round two except that the amount of money available for those choosing to dictate (that is, to opt in) was increased by various amounts, reaching \$20 in Round 6. Thus, these subsequent rounds provide an incentive for those who have opted out to opt back in. Presumably, those who have opted out are those who dislike not sharing, and hence the additional  $w$  potentially compensates them for the disutility of entering the sharing environment.

### 5.3 RESULTS

Table 2 shows that on average, about 30% is shared. When the treatment is anonymous, male sharing falls by about 15 to 30 percent whereas female sharing declines by much less.

Table 3 examines the determinants of whether or not sharing occurred in Decisions 1 and 2. Anonymity slightly reduces the odds of sharing, but this is not significant. The sorting option dramatically reduces the probability of sharing—by 35 to 55 percent!

Table 4 examines the determinants of the proportion shared of the available endowment. Note that ‘opting out’ causes this proportion to be zero. The average amount shared is about 30% in the non-sorting condition. In the sorting condition, this falls by about 15 percentage points (a reduction of half).

Table 5 shows that the size of the endowment ( $w$ ) has only a minor effect on the proportion shared.

Table 6 shows that the ‘dislike not sharing’ type—those who share in decision 1 then opt-out in decision 2—is the most common type.

Table 7 shows that the number of subjects who ‘opt-in’ is rising in the size of the endowment. Those who opt-in in subsequent rounds are those who shared more in Decision 1. In other words, the more ‘generous’ exited immediately and had to be compensated the most to re-enter the sharing condition!

Figure 1a summarizes many results. About 75% of players opt-out on Decision two. As the endowment rises to \$20, up to 2/3rds of those who opted-out opt back in. As they opt-in, the percent sharing rise, which implies that those who dislike not sharing usually share in the sharing condition. But the amount shared does not rise nearly as fast as the endowment. The average amount shared is \$3.00 in the first round and \$4.50 in the last round, despite the doubling of the endowment. This likely reflects the zeros among those not opting in. Those who opt-in do share approximately the same

*percentage* whenever they participate in the sharing condition (this is shown in Table 5).

#### 5.4 CONCLUSIONS

[quoting selectively from the article] “In the real world people regularly sort into and out of economic environments such as firms, markets, and institutions; but in the laboratory these sorting decisions are largely ignored. Instead, subjects are typically placed in one particular kind of situation and forced to make a choice that they might avoid making outside the laboratory.

When individuals are forced to play a dictator game, the majority share. But when they are allowed to opt out of the game, the majority does not share. Choosing subjects randomly, and forcing them to play the dictator game, would lead us to believe that sharing is pervasive in the world outside the laboratory. However, allowing people to avoid the sharing situation might lead to the opposite conclusion, namely that a subset of individuals share, but the majority avoid situations where sharing is possible.

Individuals may share not because they like to share, but because they dislike not sharing. In particular, some of those who appear to be most fairness-minded in the forced-choice experiments are likely to avoid environments where they can act fairly. Some individuals share in dictator games (without an outside option) not because they value implementing fair outcomes, but because they feel compelled to for some other reason. Moreover, the more such subjects feel compelled to share, the higher the price they require for entering the sharing environment. Thus, some of the people who appear the most willing to share are the least likely to enter environments where sharing is possible.”

This is an ingenious paper that will have a significant impact on (some) economists’ thinking about the external validity of laboratory experiments. Implicitly, it makes a strong case for using ‘field experiments’ instead of lab experiments to learn about the world. It also suggests that naturalistic field experiments may not be enough to gain external validity. To maximize generality, the experiment should not place individual’s in environments that they would normally self-select out of. By placing individuals in these settings, we may coerce subjects to engage in behaviors that we would rarely expect to find them performing in the outside world.

Q: Based on LMW, what are some core failures of external validity in the laboratory fairness experiments such as FFT?

6 FURTHER DOUBTS ABOUT EXTERNAL VALIDITY: “THE BEHAVIORALIST MEETS THE MARKET”  
BY LIST (2006)

In 1981, Nobel laureate George Stigler wrote that “[When] self-interest and ethical values with wide verbal allegiance are in conflict, much of the time, most of the time in fact, self-interest theory . . . will win.” Yet, this is not the finding of numerous experiments on gift exchange. The question of this study is to what extent does market discipline tend to reduce or eradicate the influence of ‘social preferences’ seen in the lab.

The innovation of the analysis (as with many studies by John List) is to combine observations from the field with observations from the lab using the same set of participants. This allows a direct test of the degree to which laboratory behavior generalizes to the field. The market that List usually works in is the sportscard trading market. This is something of an oddball environment but List appears to have great expertise here (one would not be surprised to learn that he is a sportscard trader). This is quite a complicated experiment, but also an ingenious one.

After much thought, I concluded that this experiment simply has too many moving parts to adequately summarize in class. Instead, I will discuss a much simpler List experiment below. I do highly recommend the paper, however. It’s not too complicated to read; it is simply too elaborate to cover in our brief class meeting given the other material for today’s lecture.

The bottom line of this study is this: Sophisticated sportscard dealers placed in the laboratory appear to exercise ‘social preferences’ by engaging in gift exchange. But when placed on the trading floor with real cash and merchandise, dealers generally do not reciprocate gestures that look like gift exchange. Rather, they take advantage of customers who ask for—and pay for—high quality sportscards if these customers are unable to objectively evaluate card quality.

7 A FIELD EVALUATION OF GIFT EXCHANGE: “PUTTING BEHAVIORAL ECONOMICS TO WORK:  
TESTING FOR GIFT EXCHANGE IN LABOR MARKETS USING FIELD EXPERIMENTS” BY GNEEZY  
AND LIST (2006)

This paper—not on your reading list but now posted on the 14.11 class web site—makes a very simple but nevertheless important contribution to the Gift Exchange literature by probing the external validity of gift exchange experiments. Almost all of the evidence for gift exchange (perhaps *all* of the evidence) is based on either anecdotes or lab tests like in the FFR. This paper provides a test of gift

exchange in the field.

#### 7.1 PROTOCOL 1: DATA ENTRY

Students were recruited to work computerizing the holdings of a library. Posters offered \$12 an hour for *one-time* work, implying that this would not be a repeated interaction. Each participant performed the task alone without viewing other members. In the control treatment, participants were paid \$12. In the experimental treatment, they were told after the training was complete that they would receive \$20 per hour. Output was monitored every 90 minutes.

Students were not told that they were participating in an experiment. Nineteen workers were hired for 6 hours. There were 10 controls and 9 treatment subjects.

Results of this treatment are clear cut. In the first 90 minutes, treated subjects produce about 25 percent more output per hour. In the second 90 minutes, they produce 10 percent more output. In the next two 90 minute periods, their output is almost identical to that of untreated subjects.

#### 7.2 PROTOCOL 2: FUND-RAISING

The task is door-to-door fund-raising for the Natural Hazards Mitigation Research Center. 23 fund-raisers were recruited, 13 in the no-gift condition and 10 in the gift condition. All fund-raisers were told they would be paid \$10 per hour during training and employment. After the training was completed, subjects met on a Saturday morning to begin fund-raising. The gift-treatment subjects were told at that time that they would be receiving \$20 per hour instead of \$10.

Subjects were not told that they were participating in an experiment. The gift and no-gift groups were kept separate so that there was no possibility of spillovers/contamination. Subjects were instructed to write down the amount of each contribution and time it was received so that it was possible to track output over the course of the day.

Results are again clear. In the first three hours, the gift group raises about 75% more per hour than the non-gift group: \$11.00 vs. \$6.40 per hour. In the second three hours, the gift group raises only 5% more per hour than the non-gift group: \$7.03 vs. \$6.63. Clearly, paying twice as much per hour was not a worthwhile investment.

### 7.3 INTERPRETATIONS

An obvious conclusion here is that time span matters. Gift exchange did work initially—it’s just that its effect was not durable. The authors offer one explanation for this: ‘hot’ and ‘cold’ responses. The initial ‘hot’ response is for participants to feel an upwelling of gratitude that increases motivation. The slightly longer term ‘cold’ response is for participant motivation to fall back to the ‘standard’ level, whatever that is. Why does this occur? Several possibilities:

- After the ‘hot’ response wears off, participants realize that there is no *marginal* incentive to work harder; no matter how little or much they fund raise, their pay is still \$20 per hour.
- One can also invoke Prospect Theory. When the reference pay is \$10, the \$20 rate leaves considerable psychic surplus. After a few hours, the reference rises to \$20 (participants now expect it) and so participants do not feel (much) more satisfied with their roles than they would have under the \$10 pay scheme.
- It would be fascinating to apply a subsequent treatment in which pay was *reduced* to \$10 on a second day of fund-raising. I suspect that this would have *durable* effects—treated participants would produce less output than untreated participants during both halves of the day.

Does this experiment spell the death of the theory of Gift Exchange? I very much doubt it. One can lodge several objections about external validity:

- The fact that this is a one-shot interaction seems at odds with the general gift exchange model where we believe that employers and employees continually ‘regift’ each other. Of course, one could respond that this setting has an alternative explanation: in repeated games without a finite (or known) duration, cooperation can emerge as a Nash equilibrium strategy (that is, the benefits over the long-term of cooperating outweigh the short-term gains that come from not returning the gift).
- It may be that gift exchange is more salient for negative than positive reciprocity. As in my conjecture above, removing the gift may do more harm than adding the gift yielded benefits.
- One can also invoke the LMW sorting critique. The subjects recruited for this study were not selected for reciprocity. If an employer is going to use a gift exchange wage policy, it will

want to select workers who tend to respond cooperatively. Hence, gift exchange that operates within competitive markets might only take place with selected participants. Case in point, it's unlikely that an employer would, for no announced reason, double the bargained wage prior to work beginning. More artificial still in this case is that the employer does not announce that along with a doubling of the wage comes an expectation (or request or stated desire) for higher output. Thus, it's not entirely clear what the gift is supposed to stimulate in this setting. That said, output did rise initially, so there clearly is a reciprocation reflex, though it does not seem to be long-lasting.