Behavioural Economics:
Theory and Evidence on Bargaining

University of Oxford, Michaelmas Term 2013

and

Asia Summer Institute in Behavioral Economics
National University of Singapore, 22 July-2 August 2013

Vincent P. Crawford, University of Oxford, All Souls College,
and University of California, San Diego

(with thanks to Sergei Izmalkov, Paul Milgrom,
Al Roth and Muhamet Yildiz)

Revised 3 July 2013

Bargaining

Bargaining is at the heart of game-theoretic microeconomics.

There are two prominent strands of theory, and some important experiments that raise questions about the theory, still unresolved.

The theory includes the unstructured/cooperative theory of Nash (1950 *Eonometrica*), the structured/cooperative theory of Nash (1953 *Econometrica*) and the structured/noncooperative theory of Rubinstein (1982 *Econometrica*).

The experiments include the unstructured bargaining experiments of Roth, Murnighan, and collaborators, culminating in Roth and Murnighan (1982 *Econometrica*), and structured bargaining experiments.

In these slides I briefly consider both theory and experiments.

Rubinstein's structured/noncooperative theory of bargaining

Rubinstein's theory postulates that bargaining proceeds via a given noncooperative game, in which two bargainers make alternating offers to each other on a fixed schedule until they reach an agreement.

In Rubinstein's version of the model there is no time limit.

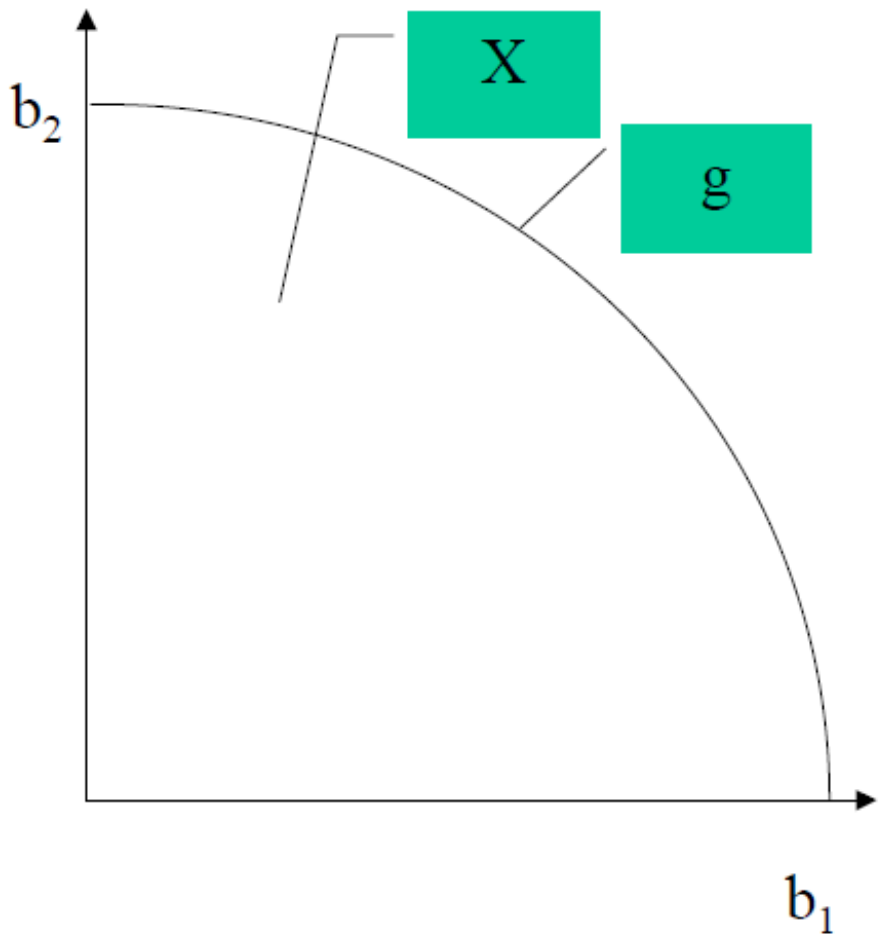
(Ingolf Stahl earlier studied a finite-horizon version, of which an extreme case is the one-period ultimatum game.)

Delay is assumed to be costly (except in the ultimatum game, where delayed agreements are impossible anyway).

The model allows the agreement to be about anything or things that are continuously variable, but I assume for simplicity that the bargainers are bargaining over the division of a dollar.

But bargainers are allowed to have nonlinear vN-M utility functions over money, making divide-the-dollar like the general case.

[Snapshots from MIT open courseware notes by Izmalkov and Yildiz]



- $N = \{1,2\}$
- X = feasible expected-utility pairs $(x,y \in X)$
- $U_i(x,t) = \delta_i^t x_i$
- $D = (0,0) \in X$ disagreement payoffs
- g is concave, continuous, and strictly decreasing

$T = \{0,1,\ldots, t,\ldots\}$

At each t, if t is even,

- Player 1 offers some x
- Player 2 Accepts or Rejects the offer
- If the offer is Accepted, the game ends yielding x
- Otherwise, we proceed to t + 1

if t is odd,

- Player 2 offers some y
- Player 2 Accepts or Rejects the offer
- If the offer is Accepted, the game ends yielding y
- Otherwise, we proceed to t + 1

(Recall that the outcome is described as a pair of utilities.)

To build intuition, consider the one-period version, the ultimatum game.

It has a unique subgame-perfect equilibrium in which player 1 (the "proposer") makes a proposal in which he gets all of the surplus and player 2 (the "responder") accepts (despite indifference; explain as limit).

Because the proposer gets all the surplus, he has an incentive to structure his proposal to maximize it, so the outcome is Pareto-efficient.

(In divide-the-dollar this means only not wasting money; but efficiency holds in much more general models of bargaining over contracts.)

Thus, the "noncooperative" contracting game yields a "cooperative" equilibrium outcome.

The model *explains* the efficiency of the bargaining outcome as the outcome of individually rational bargaining, rather than just assuming it.

This is important because it allows analysis of how bargaining institutions and information conditions affect efficiency: but that's another story.

Now consider a finite-horizon version of the model due to Ingolf Stahl.

Again there is a unique subgame-perfect equilibrium, in which player 1 makes a proposal in which he gets all of the surplus from reaching an agreement immediately relative to a one-period delay (anticipating subgame-perfect equilibrium in the subgame that would follow rejection), and player 2 immediately accepts (again despite indifference).

The subgame-perfect equilibrium outcome is again Pareto-efficient, because player 1 has an incentive to make a proposal that maximizes the surplus and there's no delay.

Surplus-sharing is entirely determined by delay costs, with a first-mover advantage for player 1 when there is an odd number of periods, which goes away as the discount factor $\delta \rightarrow 1$ and the horizon $\rightarrow \infty$.

In the limit, with equal discount factors, surplus-sharing approaches the Nash (1950) bargaining solution, discussed below. In divide-the-dollar with linear utility functions, this reduces to an equal split.

Now consider Rubinstein's (1982) infinite-horizon version of the model.

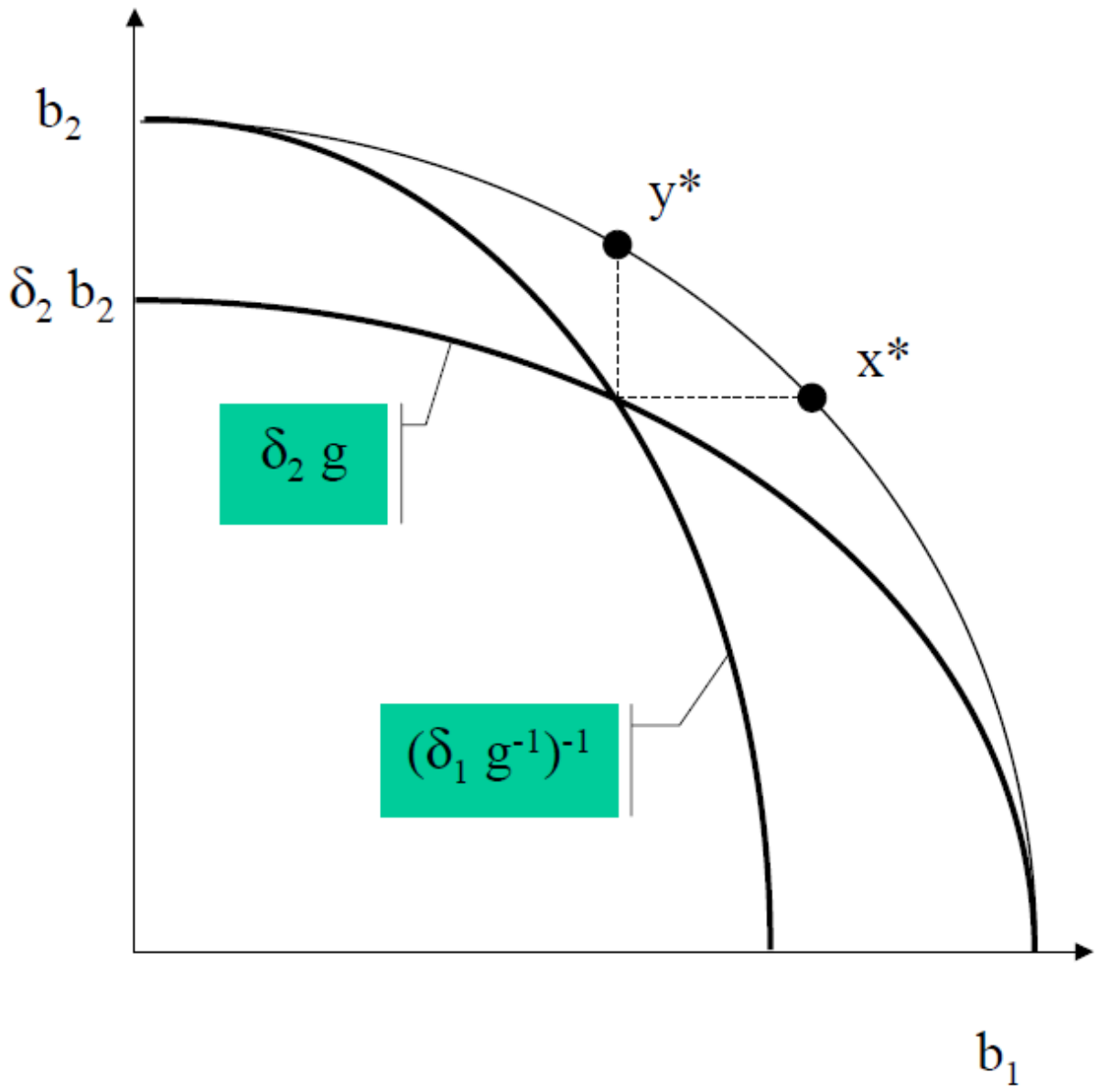
Once again there is a unique subgame-perfect equilibrium in which player 1 makes a proposal in which he gets all of the surplus from reaching an agreement immediately relative to a one-period delay (anticipating subgame-perfect equilibrium in the subgame that would follow rejection), and player 2 immediately accepts (despite indifference).

Uniqueness was trivial before, but it's nontrivial now, and a fragile result.

The subgame-perfect equilibrium outcome is again Pareto-efficient, for the same reasons; and it equals the limit of the finite-horizon subgame-perfect equilibria as the horizon approaches infinity.

Surplus-sharing is still determined by delay costs, with a first-mover advantage for player 1, which goes away as the discount factor $\delta \rightarrow 1$.

With equal discount factors surplus-sharing again approaches the Nash solution, or equal split in divide-the-dollar with linear utility functions.


b_2
$\delta_2\ b_2$
y*
x*
$\delta_2\ g$
$(\delta_1\ g^{-1})^{-1}$
b_1

Theorem:

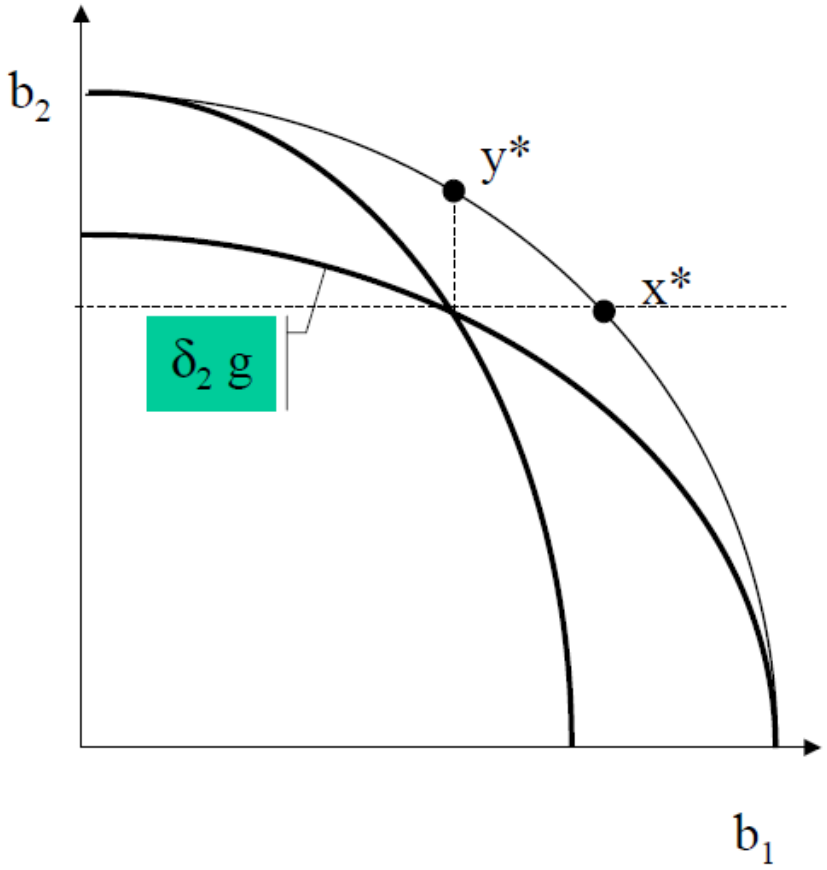
The following is the unique subgame-perfect equilibrium:

- player 1 always offers x^*;
- player 2 accepts an offer x iff $x_2 \geq x^*_2$;
- player 2 always offers y^*;
- player 1 accepts an offer y iff $y_1 \geq y^*_1$;

When the utility functions are linear as in divide the dollar, for instance, player 1 initially offers $x^* = [(1-\delta_2)/(1-\delta_1\delta_2), \delta_2(1-\delta_2)/(1-\delta_1\delta_2)]$ and player 2 immediately accepts it, so there is no delay.

When $\delta_1 = \delta_2 = \delta$, this reduces to $x^* = [1/(1+\delta), \delta/(1+\delta)]$, and as $\delta \rightarrow 1$, $x^* \rightarrow [1/2, 1/2]$.

Proof (it is a SPE)



Use single deviation principle:

1. If player 2 rejects an offer x at t, she will get y_2^* at t+1. Hence, Accept iff $x_2 \geq \delta_2 y^*_2 = x^*_2$ is optimal at t.
2. At t, it is optimal for 1 to offer

$x^* = \text{argmax } \{x_1 | x_2 \geq x^*_2\}$.

7

Nash's cooperative theory of bargaining

Nash's cooperative bargaining solution is perhaps the leading model of bargaining in economics.

He assumed that two bargainers are faced with a set of feasible agreements. If they can agree on one, it will be the outcome. If not, the outcome will be an exogenously given disagreement outcome.

He also assumed that the bargainers have vN-M utility functions defined over the feasible agreements and the disagreement outcome, and took the resulting utility-possibility set and disagreement utilities as the data of the bargaining problem, assuming that both are common knowledge.

Nash sought to axiomatize a reasonable bargaining solution (he was not clear whether it was intended to be positive or normative), which mapped these data into a "solution" giving each bargainer a unique vN-M utility.

[Snapshots from MIT open courseware notes by Milgrom and Yildiz]

Nash Bargaining Problem

- $N = \{1,2\}$ – the agents
- $S \subset R^N$ -- the set of feasible expected-utility pairs
- $d = (d_1,d_2) \in S$ – the disagreement payoffs
- A *bargaining problem* is any (S,d) where
  - S is compact and convex, and
  - $\exists x \in S$ s.t. $x_1 > d_1$ and $x_2 > d_2$.
- B is the set of all bargaining problems.
- A *bargaining solution* is any function

  $f : B \to R^N$ s.t. $f(S,d) \in S$ for each (S,d).

Nash Axioms

1. **Expected-utility Axiom [EU]** (invariance under affine transformations): $\forall(S,d), \forall(S',d'), a_i > 0$

$$\left.\begin{array}{l} S' = \{s' \mid s'_i = a_i s_i + b_i \ \forall i \in N\} \\ d'_i = a_i d_i + b_i \ \forall i \in N \end{array}\right\} \Rightarrow f_i(S',d') = a_i f_i(S,d) + b_i \ \forall i \in N$$

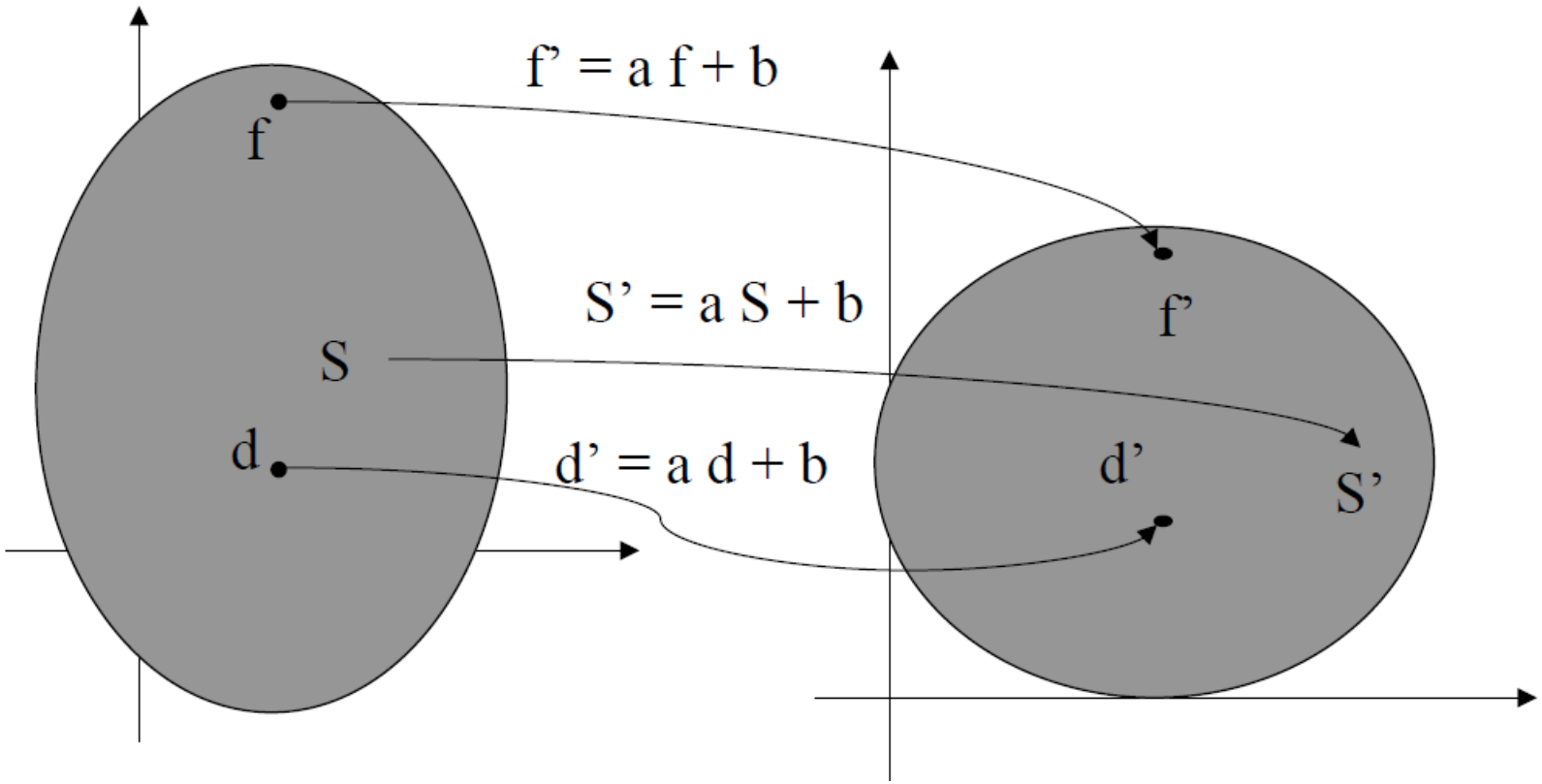
2. **Symmetry [Sy]:** Let (S,d) be symmetric: $d_1 = d_2$ and [$(x_1,x_2) \in S$ iff $(x_2,x_1) \in S$]. Then,

$$f_1(S,d) = f_2(S,d).$$

3. **Independence of Irrelevant alternatives [IIA]**: if $T \subset S$ and $f(S,d) \in T$, then $f(T,d) = f(S,d)$.

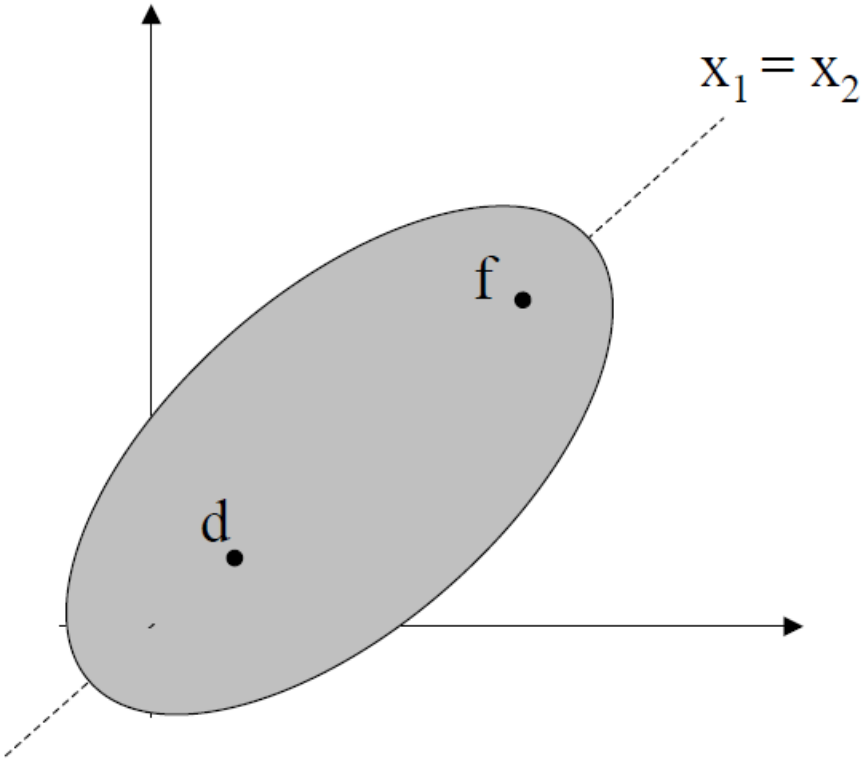
4. **Pareto – Optimality [PO]:** if $x,y \in S$ and $y > x$, then $f(S,d) \neq x$.

Expected-utility Axiom

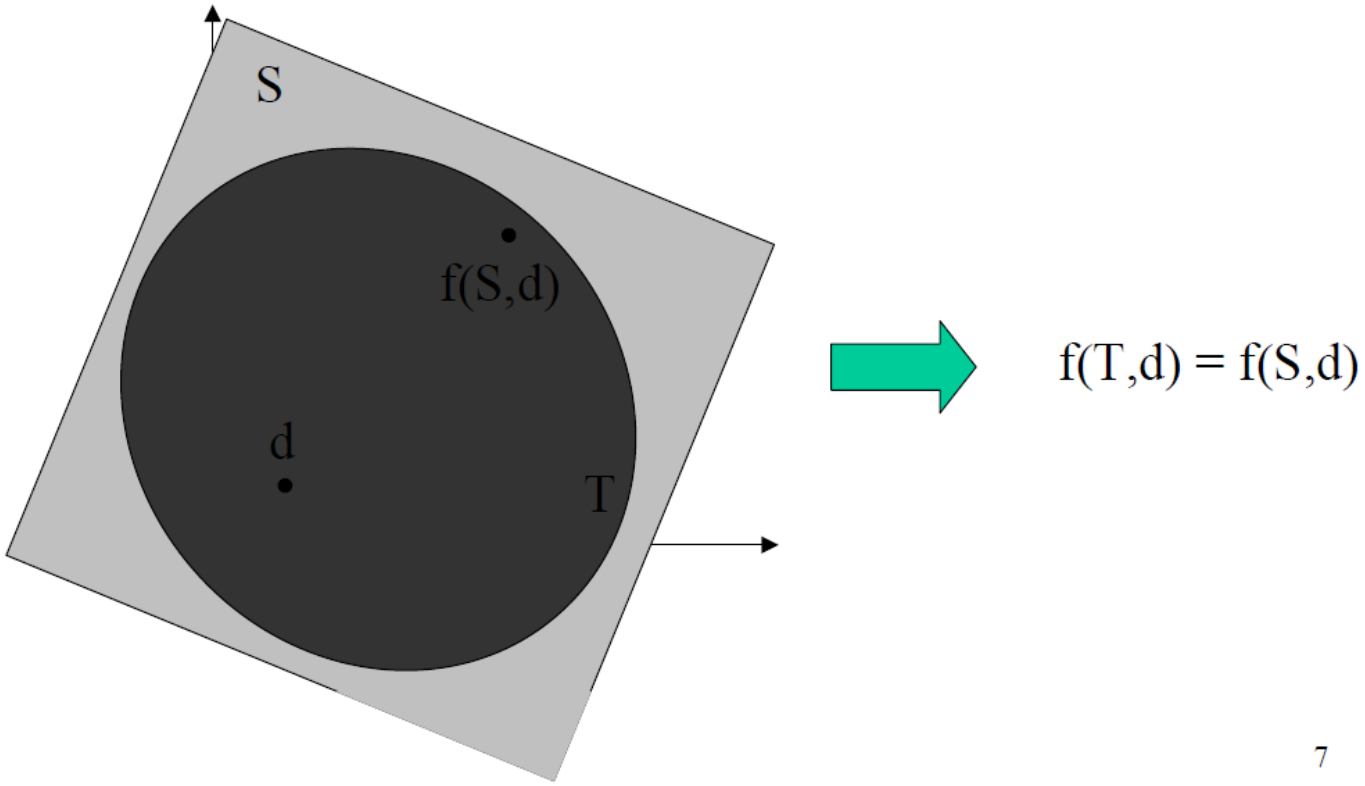


This is not really an implication of rationality and it destroys the language in which interpersonal comparisons would have to be expressed.

Symmetry



Independence of irrelevant alternatives



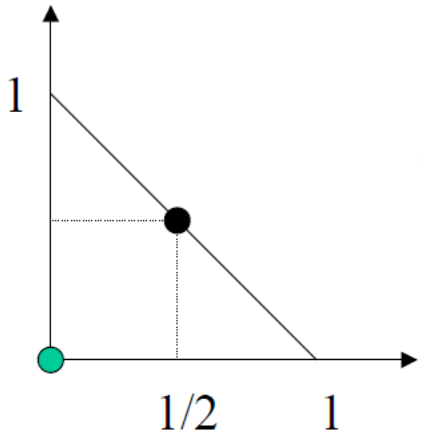
7

Note: Natural for individual choice, but less compelling for bargaining.

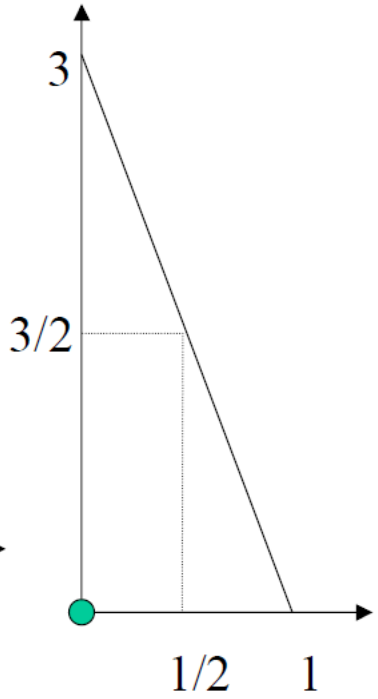
Nash Bargaining Solution

$$f^*(S,d) = \arg\max_{\substack{s \equiv (s_1, s_2) \in S \\ s > d}} (s_1 - d_1)(s_2 - d_2).$$

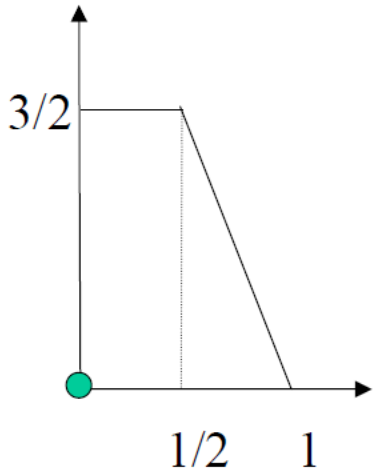
Examples



SY and PO



EU



IIA

Nash's Theorem

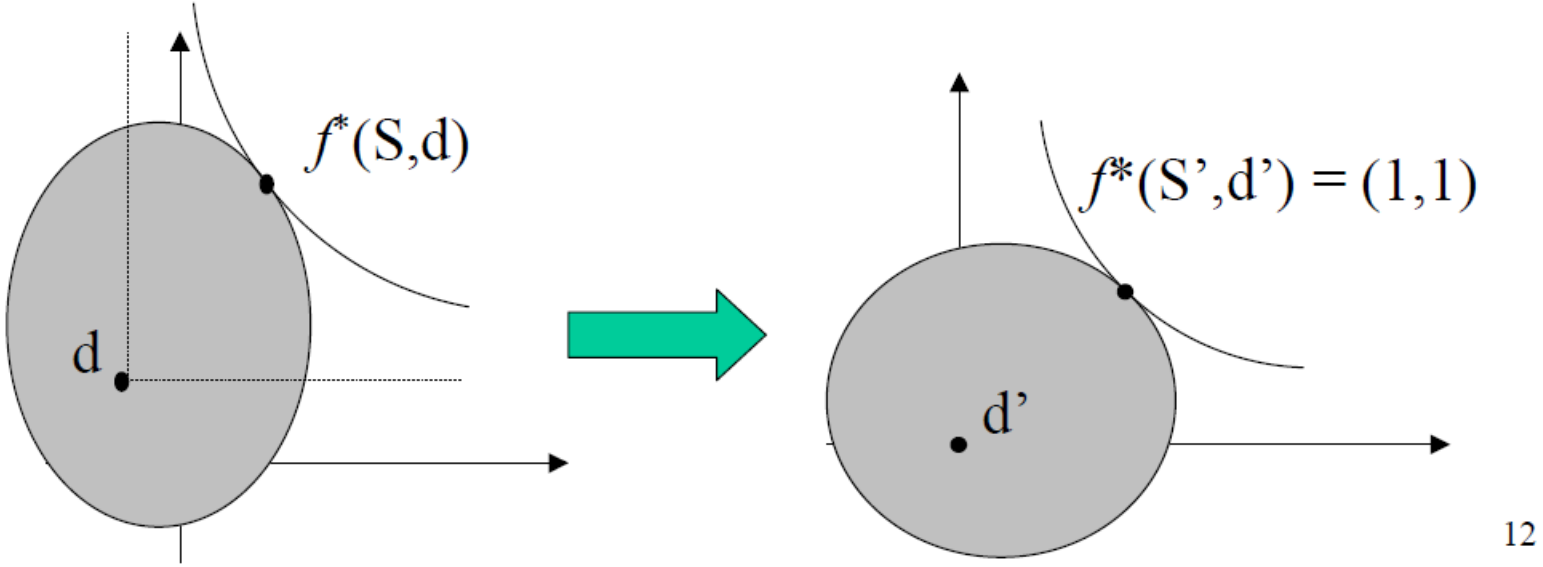
Theorem: A bargaining solution f satisfies the Nash Axioms (EU,Sy,IIA,PO) if and only if

$$f = f^*.$$

$$f^*(S,d) = \arg\max_{\substack{s \equiv (s_1,s_2) \in S \\ s > d}} (s_1 - d_1)(s_2 - d_2).$$

Proof of Nash's Theorem

1. Check: f^* satisfies the Nash axioms. (easy)
2. Take any (S,d). Transform it to (S',d') so that d' = 0, and f^*(S',d') = (1,1). Under [Sy,IIA,PO], f(S',d') = f^*(S',d') = (1,1). &EU $\Rightarrow$ f(S,d) = f^*(S,d). QED



12

Nash Axioms

1. **Expected-utility Axiom [EU]** (invariance under affine transformations): $\forall(S,d), \forall(S',d'), a_i > 0$

$$\left.\begin{array}{l} S' = \{s' \mid s'_i = a_i s_i + b_i \ \forall i \in N\} \\ d'_i = a_i d_i + b_i \ \forall i \in N \end{array}\right\} \Rightarrow f_i(S',d') = a_i f_i(S,d) + b_i \ \forall i \in N$$

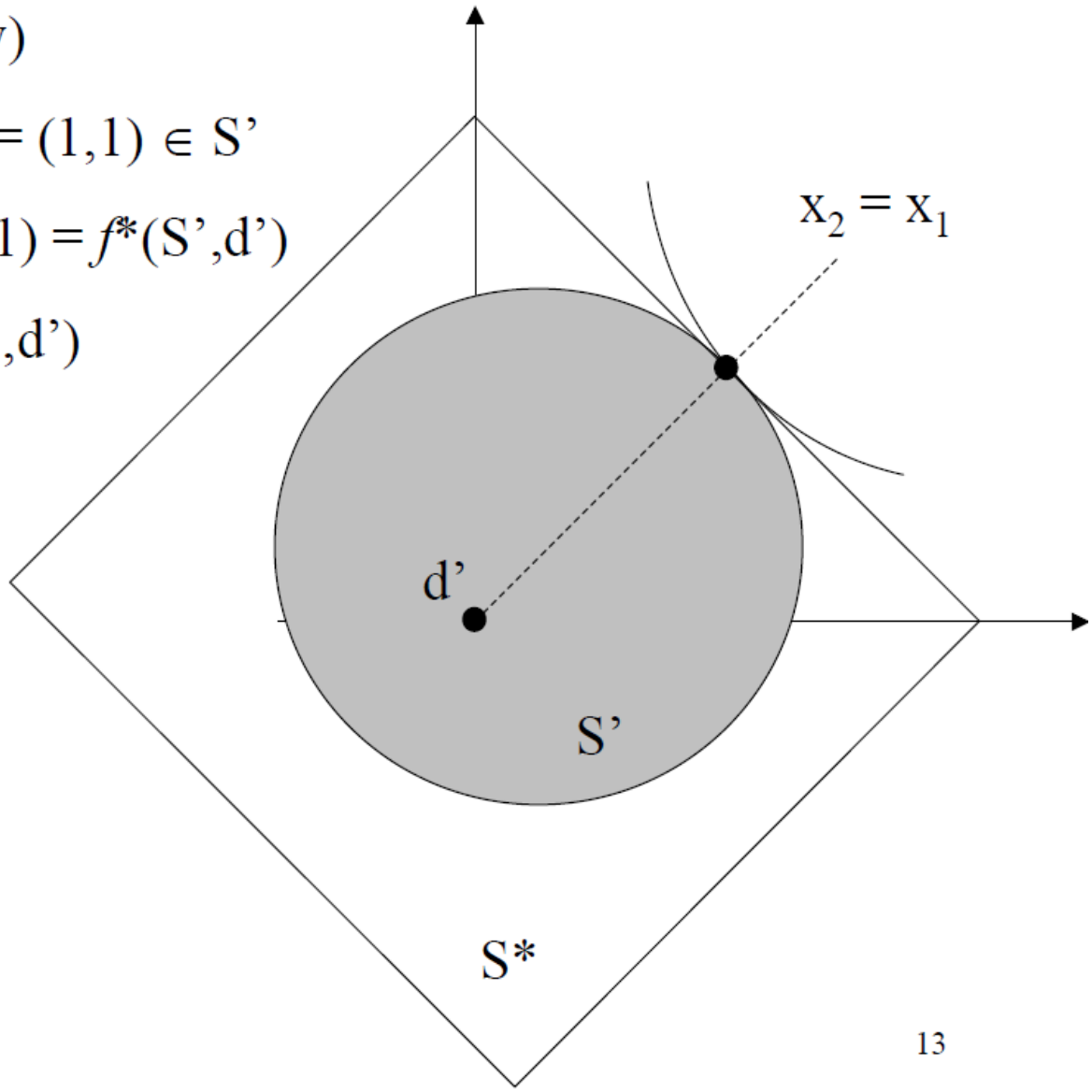
2. **Symmetry [Sy]:** Let (S,d) be symmetric: $d_1 = d_2$ and [$(x_1,x_2) \in S$ iff $(x_2,x_1) \in S$]. Then,

$$f_1(S,d) = f_2(S,d).$$

3. **Independence of Irrelevant alternatives [IIA]**: if $T \subset S$ and $f(S,d) \in T$, then $f(T,d) = f(S,d)$.

4. **Pareto – Optimality [PO]:** if $x,y \in S$ and $y > x$, then $f(S,d) \neq x$.

- S* is symmetric. (how)
- &Sy&PO $\Rightarrow f(S^*,d') = (1,1) \in S'$
- &IIA $\Rightarrow f(S',d') = (1,1) = f^*(S',d')$
- &EU $\Rightarrow f(S,d) = f^*(S',d')$



13

Notes: It's always possible to do the transformation in Step 2 because the EU axiom yields four degrees of freedom.

The transformed utility-possibility set always lies inside the triangle from the origin to $x_1 + x_2 = 1$ because if not, one could reach a higher Nash product in the rescaled problem.

As the proof makes clear, Nash's solution can be viewed as a generalization of the equal-gains-over-disagreement solution to nonlinear utility-possibility frontiers.

It is a remarkable coincidence that the limiting subgame-perfect equilibrium of Rubinstein's noncooperative bargaining model and Nash's cooperative bargaining solution yield identical outcomes.

Although not demonstrated here, this equivalence extends beyond divide the dollar to nonlinear utility-possibility frontiers.

Nash (1953 *Econometrica*) further supported his cooperative solution, beginning what is now called the “Nash program,” which seeks to unify the cooperative and noncooperative approaches to game theory.

He introduced what is now called the “Nash demand game”, in which bargainers simultaneously make proposals, x and y (pairs of utilities).

If their proposals consistent with some allocation in the utility possibility set, they are implemented. If not, they get the disagreement outcome.

This game has a continuum of pure-strategy equilibria, which efficiently share the surplus and are the utility counterpart of Edgeworth’s contract curve. (It also has a continuum of inefficient mixed-strategy equilibria.)

Nash then gave a “smoothing” argument, foreshadowing trembling-hand perfection, which singled out his cooperative solution.

More on this below.

Experimental tests

One difficulty in testing theories of bargaining is that their predictions are sensitive to bargainers' vN-M utility functions, which are unobservable.

(Practically speaking, we could assume approximate risk-neutrality; but this is would be a weakness testing a theory that gives a central role to nonlinearity of vN-M utility functions, and would not convince theorists.)

Another difficulty is that bargaining in the world is unstructured, without the cut-and-dried rules of noncooperative models like Rubinstein's.

Arguably, the best chance for a cooperative solution like Nash's to describe behavior is when bargaining is unstructured (despite Rubinstein's link between structured bargaining and Nash's solution).

Both difficulties were overcome in a remarkable series of experiments on unstructured bargaining by Roth and his collaborators during the late 1970s and early 1980s (well summarized in Roth 1985, 1987).

Their unstructured designs allow more informative tests of theories of bargaining than is possible by imposing a structure.

The rules of bargaining were as follows:

If subjects could agree how to share the lottery tickets by an announced deadline the agreement was enforced; otherwise they got nothing.

Subjects could make any binding proposal they wished, or accept their partner's latest proposal, at any time.

They could also send nonbinding messages at any time, except that they could not identify themselves or, in some treatments, reveal their prizes.

The environment was public knowledge, with exceptions noted below.

Specifically (in Roth's description):

Each participant sat at a visually isolated terminal of a networked computer laboratory. Participants could send each other text messages (which passed through a monitor's terminal) containing anything other than information about personal identity (e.g. "I am sitting in station 24 of the foreign language building, wearing a blue windbreaker" was not allowed).

They could also send numerical proposals.

After a message or a proposal was entered, it appeared on the screen with a prompt asking whether you wanted to edit it or transmit it to your bargaining partner. In order to accept a numerical proposal, you had to transmit the identical proposal back. (e.g. my share = 67%, your share = 33%).

There was a fixed time period, and a clock on the screen counted off the time. If agreement on a numerical proposal had not been reached by the end of the time period, the game ended with disagreement.

The designs controlled bargainers' unobservable vN-M utility functions via the binary lottery procedure of Roth and Malouf (1979).

Specifically, pairs of subjects bargained over a fixed total of 100 lottery tickets, with each subject's share determining his probability of winning the larger of two possible monetary prizes, specific to him.

If subjects preferred higher probability of more money to lower, as implied by EU preferences and many others, then the binary lottery procedure makes them risk-neutral in lottery tickets, always preferring more to less. (We are all supposed to be risk-neutral in probabilities.)

Arguably, with the rest of the structure publicly announced except for the sizes of subjects' prizes and their information about prizes, which were sometimes withheld, this made subjects' preferences over the lottery tickets in which bargaining was conducted public knowledge.

Thus, it makes the utility-possibility set and disagreement outcome observable and allows direct tests of theories that assume common knowledge of the structure of the bargaining game.

The designs exploit invariances created by the binary lottery procedure to test both cooperative and noncooperative theories of bargaining.

Under standard assumptions the number of lottery tickets a bargainer obtains can be taken as his vN-M utility or payoff.

This makes bargaining over a fixed total of lottery tickets equivalent to a complete-information divide the dollar game with risk-neutral players, whose symmetry leads cooperative theories to predict equal division of the lottery tickets.

These conclusions are independent of players' risk preferences, prizes, or information about prizes, so that cooperative theories can be tested by observing the effects of varying those factors.

Although noncooperative theories are harder to test this way because their predictions may depend on the details of the structure, the binary lottery procedure also makes it possible to create invariances that allow such tests, as explained below.

Each treatment paired a subject whose prize was low (typically $5) with one whose prize was high ($20). A subject always knew his own prize.

The first experiment compared two information conditions: “full,” in which a subject also knew his partner's prize; and “partial,” in which a subject knew only his own prize.

The second experiment created a richer set of information conditions using an intermediate commodity, chips, which subjects could later exchange for money in private. A subject always knew his own chip prize and its value in money.

There were three information conditions: “high,” in which a subject also knew his partner’s chip prize and its value; “intermediate,” in which a subject knew his partner’s chip prize but not its value; and “low,” in which a subject knew neither his partner’s chip prize nor its value.

Subjects were prevented from communicating the missing information, and the information condition was public knowledge.

Partial and low information induce games with identical structures, given that players cannot send messages about chip or money prizes, because their strategy spaces are isomorphic (with chips in the latter treatment playing the role of money in the former) and isomorphic strategy combinations yield identical payoffs (in lottery tickets).

For the same reasons, full and intermediate information also induce games with identical structures, given that players in the latter cannot send messages about money prizes.

Any structural theory, cooperative or noncooperative, predicts identical outcomes in these pairs of treatments.

(One could define a noncooperative structural theory that allows factors like those studied in the experiments to influence bargaining, but these are conventionally ruled out in noncooperative game theory.)

A third experiment explored the strategic use of private information by giving subjects the option of communicating missing information about prizes. There were no chips, and a subject always knew his own money prize. There were four basic information conditions:

(i) neither subject knew both prizes

(ii) only the subject whose prize was $20 knew both prizes

(iii) only the subject whose prize was $5 knew both prizes

(iv) both subjects knew both prizes

Some treatments made the information condition public knowledge, while in others subjects were told only that their partners might or might not know what information they had.

Thus there were eight information conditions in all.

I first describe the observed patterns of agreements, and then discuss disagreements.

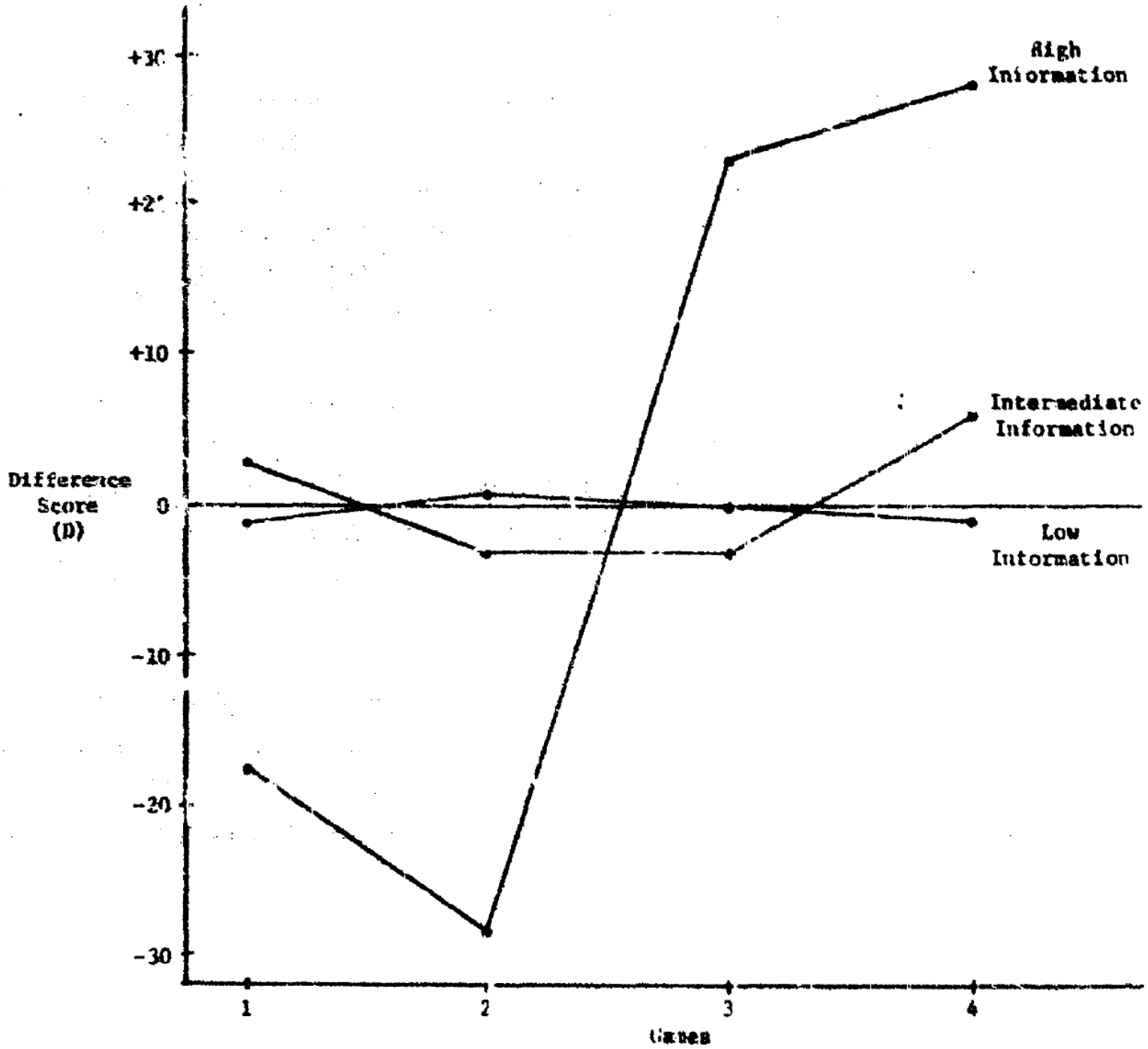
With partial information almost all subjects agreed on a 50-50 division of the lottery tickets.

With full information, the low-prize subject often asked for, and got, more than half of the lottery tickets, and agreements averaged about halfway between 50-50 and equal expected money winnings, with much higher variance: An unpredicted effect of information about the prize values.

With low and high information, respectively, agreements averaged close to 50-50 and roughly halfway between 50-50 and equal expected money winnings, again with higher variance.

With intermediate information, agreements averaged close to 50-50.

For example:


+30
+20
+10
0
-10
-20
-30
Difference Score (D)
High Information
Intermediate Information
Low Information
1
2
3
4
Games

Thus partial and low information yielded similar outcomes.

But with full and intermediate information, strategically equivalent information about money and chips affected the outcomes in very different ways, which are inconsistent with any structural theory.

The authors attributed the strong influence of subjects' prizes and information about prizes, irrelevant in traditional analyses, to the different meanings subjects assigned to chips and money outside the laboratory.

Their agreements can be summarized by postulating a commonly understood hierarchy of contextual equal-sharing norms in which subjects implemented the most "relevant" norm their public knowledge allowed, with money most relevant, then lottery tickets, and then chips.

In the third experiment agreements were largely determined by whether the $5 subject knew both prizes, clustering around 50-50 when he did not, and shifting more than halfway toward equal expected money winnings when he did.

In effect these agreements were determined by the most relevant norm in the above hierarchy that subjects could implement, using their public knowledge plus whatever private information they had incentives to reveal, on the anticipation that it would be used this way.

Subjects' revelation decisions were approximately in equilibrium in beliefs in a restricted game, in which they could either reveal the truth or nothing at all, when their beliefs are estimated from the mean payoffs in related treatments.

Frequency of Agreements in Terms of the Percentage of Lottery Tickets Obtained by the $20 Player

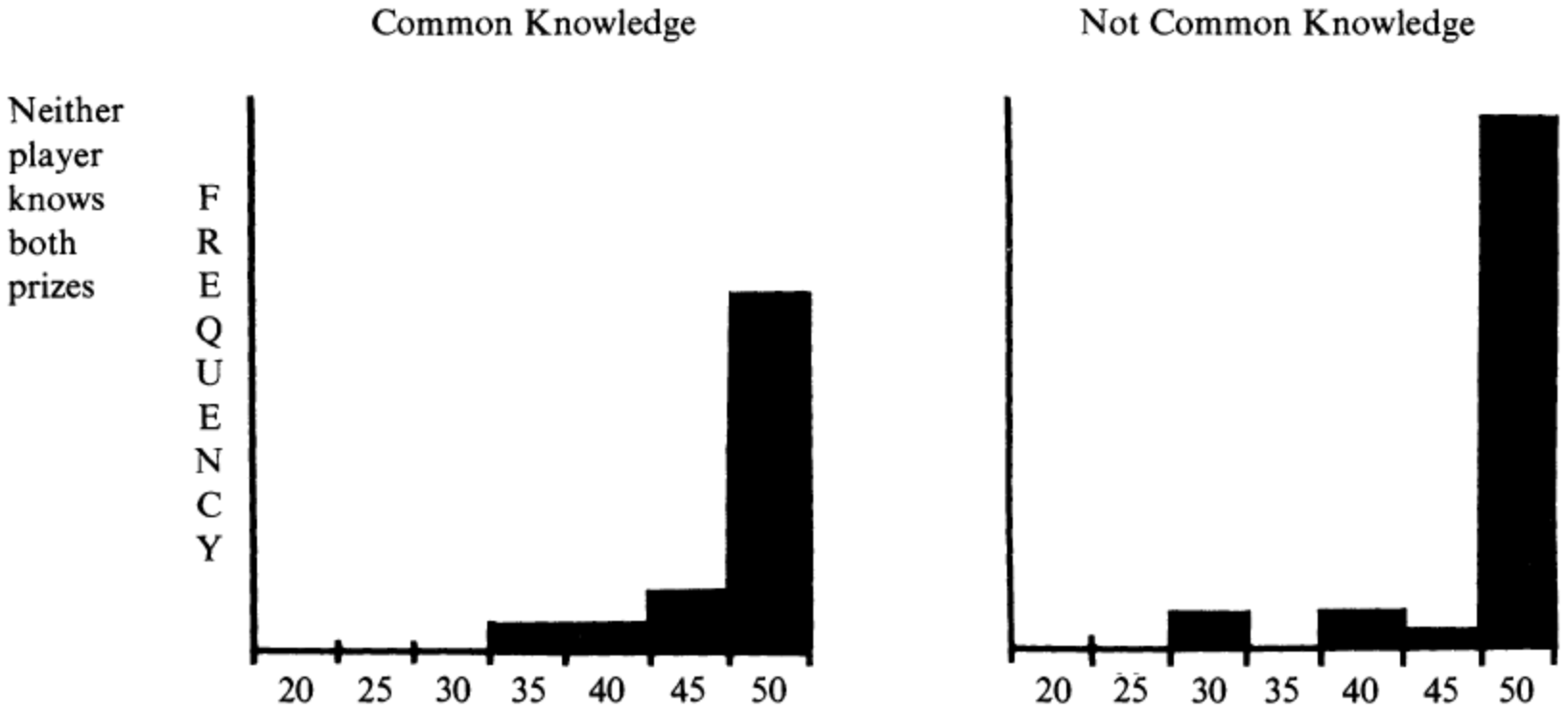


FIGURE 1.

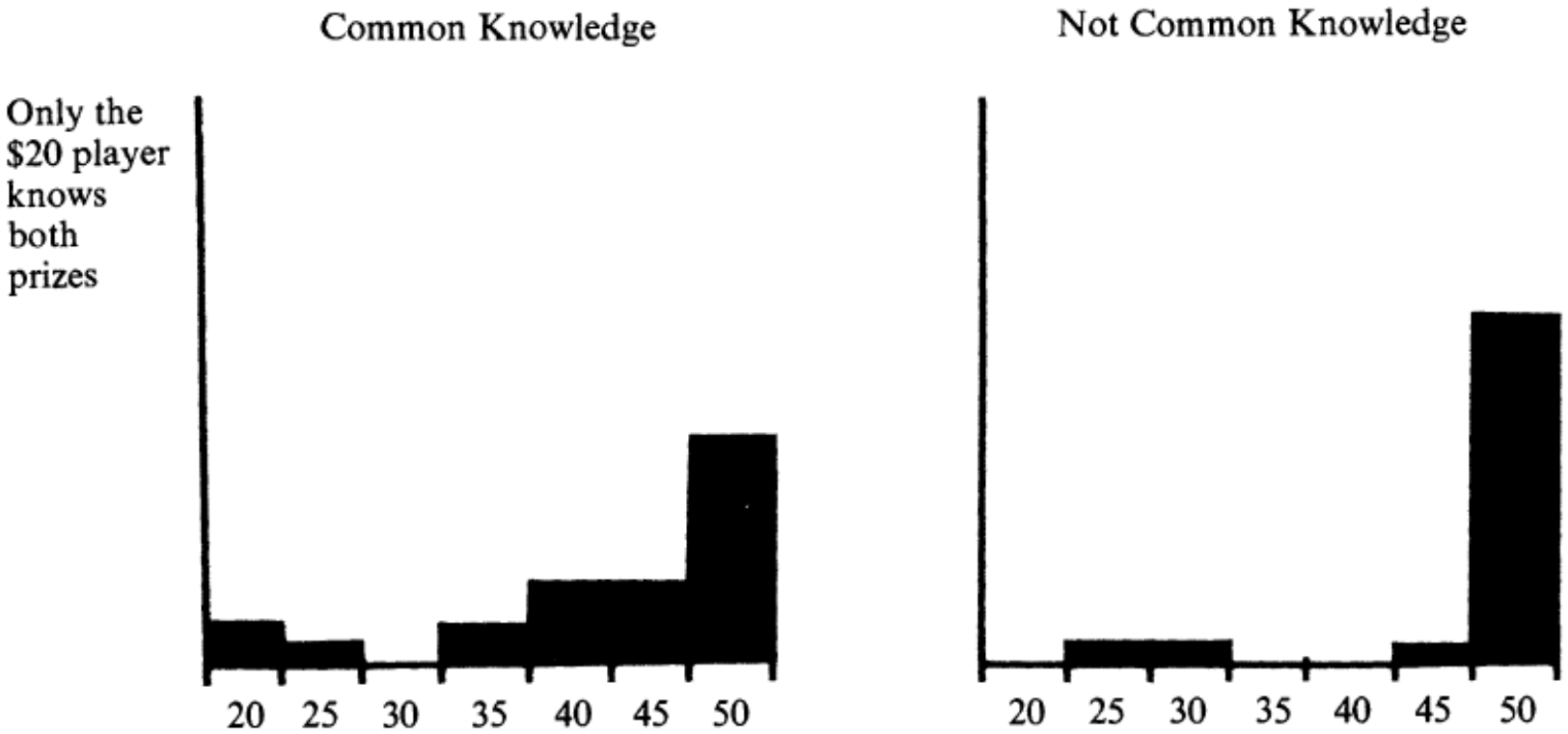

Common Knowledge
Not Common Knowledge
Only the $20 player knows both prizes
20 25 30 35 40 45 50
20 25 30 35 40 45 50

FIGURE 2.

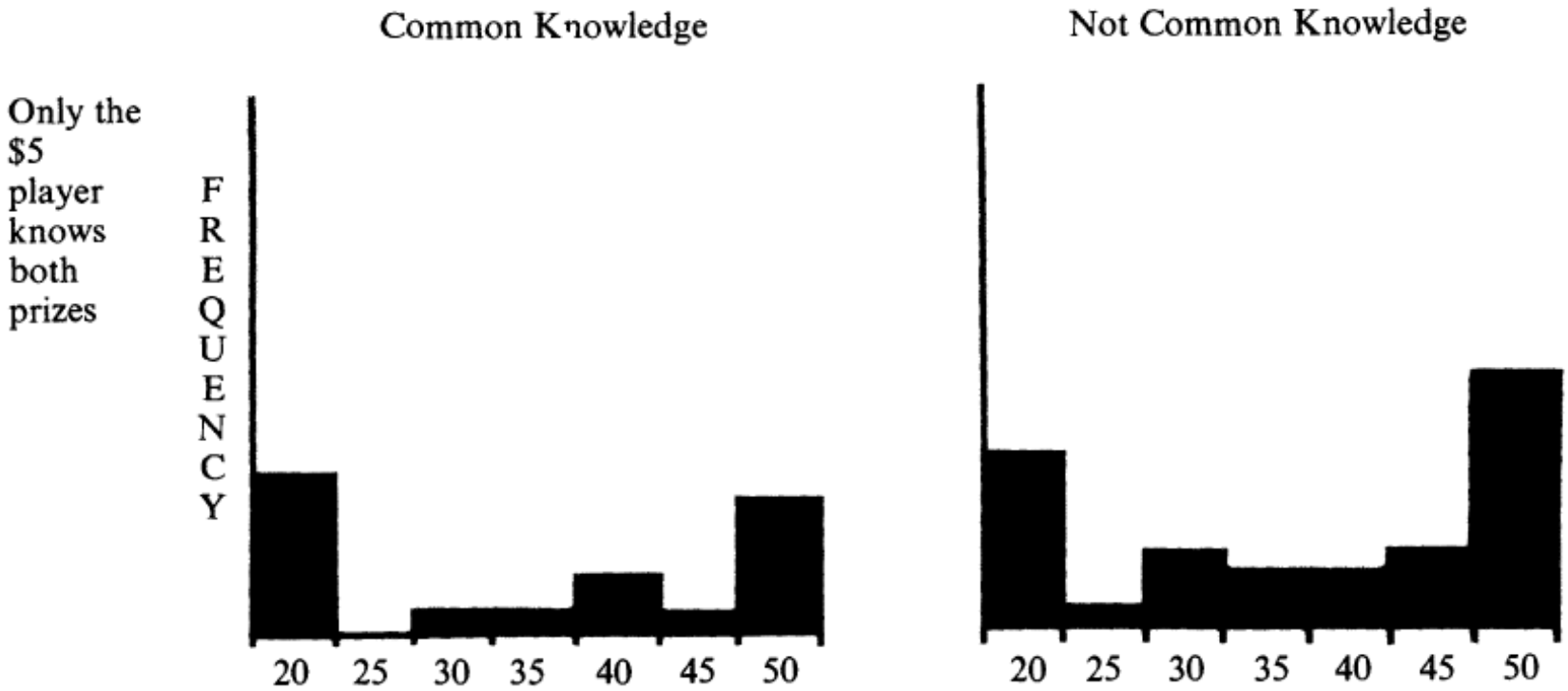

Common Knowledge
Not Common Knowledge
Only the $5 player knows both prizes
FREQUENCY
20 25 30 35 40 45 50
20 25 30 35 40 45 50

FIGURE 3.

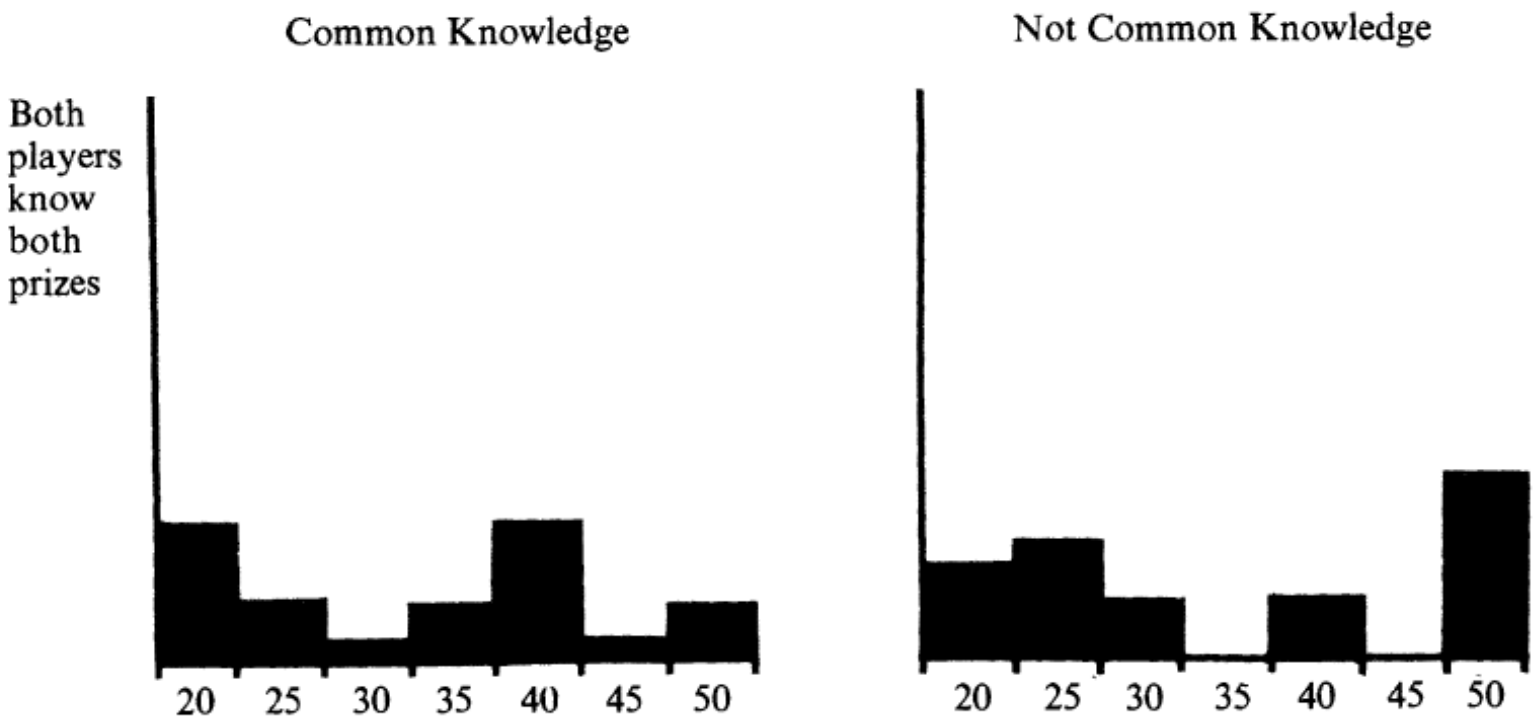

Common Knowledge
Not Common Knowledge
Both players know both prizes
20 25 30 35 40 45 50
20 25 30 35 40 45 50

FIGURE 4.

There was a subtle interplay between the use of norms and the revelation of private information.

In the public-knowledge version of condition (ii) in the third experiment, for instance, the $5 subject knew that his partner knew which agreement gave them equal expected money winnings, but the $20 subject usually refused to reveal his prize.

This left the 50-50 division the only norm that could be implemented using public knowledge.

Although many $5 subjects voiced suspicions (in transcripts) that they were being treated unfairly, in the end most settled for the 50-50 division.

The manipulation of norms by withholding information is inconsistent with nonstrategic explanations in which subjects “try to be fair”.

But most of the results can be understood using a simple strategic model, with shared ideas about fairness as coordinating principles.

This model summarizes the strategic possibilities of unstructured bargaining using Nash’s (1953 *Econometrica*) demand game, in which players make simultaneous demands, in this case for lottery tickets.

If their demands are feasible they yield a binding agreement; if not there is disagreement.

Any pair of demands that leads to an outcome that is at least as good as disagreement for each player and is Pareto-efficient is in equilibrium.

A player who reduced his demand, starting from such a pair, would lower his payoff with no compensating benefit; and a player who increased his demand would cause a disagreement, again lowering his payoff.

To see how this simple, static game can describe the complex dynamics of unstructured bargaining, assume that delay costs are negligible before the deadline, so that the timing of an agreement is irrelevant.

(This is a good approximation for the experiments and many applications to bargaining in the field.)

Then, if equilibrium is assumed, all that matters about a player's strategy is the lowest share it can be induced to accept by the deadline.

These lowest shares determine the outcome like players' demands in the demand game.

Although there are normally many efficient agreements that are better than disagreement for both bargainers, all are consistent with equilibrium in the demand game, which is thus no help in choosing among them.

Bargaining therefore generates a great deal of *strategic* uncertainty about how players will respond to its multiplicity of equilibria, even when there is no other uncertainty in the bargaining environment.

Unless the bargainers find a way to resolve this uncertainty, they may not realize any of the gains from reaching an agreement:

At the heart of the bargaining problem is a coordination problem, with players' beliefs the dominant influence on outcomes.

And bargaining remains a problem even when bargainers are fully informed about the bargaining problem.

In Schelling's words (*The Strategy of Conflict*, p. 70):

Most bargaining situations ultimately involve some range of possible outcomes within which each party would rather make a concession than fail to reach agreement at all. In such a situation any potential outcome is one from which at least one of the parties, and probably both, would have been willing to retreat for the sake of agreement, and very often the other party knows it. Any potential outcome is therefore one that either party could have improved by insisting; yet he may have no basis for insisting, since the other knows or suspects that he would rather concede than do without agreement. Each party's strategy is guided mainly by what he expects the other to accept or insist on; yet each knows that the other is guided by reciprocal thoughts. The final outcome must be a point from which neither expects the other to retreat; yet the main ingredient of this expectation is what one thinks the other expects the first to expect, and so on. Somehow, out of this fluid and indeterminate situation that seemingly provides no logical reason for anybody to expect anything except what he expects to be expected to expect, a decision is reached. These infinitely reflexive expectations must somehow converge on a single point, at which each expects the other not to expect to be expected to retreat.

In the complete model of the use of norms and the revelation of private information in the third experiment, players first decide simultaneously how much private information to reveal.

They then bargain, with ultimate acceptance decisions described by the demand game, in which there is effectively complete information.

Among the continuum of efficient equilibria in the demand game, players' beliefs are focused (if at all) by the most relevant norm their public knowledge (including any revealed private information) allows them to implement, with money most relevant, then lottery tickets, then chips.

Pure-strategy equilibria, selected this way, yield agreements that closely resemble those observed in the various treatments.

From this point of view, it is the desire to avoid a risk of disagreement due to coordination failure that explains $5 subjects' willingness to settle on the "unfair" 50-50 division in condition (ii) of the third experiment, a phenomenon that is difficult to explain any other way.

In all three experiments disagreements occurred, with frequencies ranging from 8-33%.

Disagreements were most common when both subjects knew enough to implement more than one norm, or when the information condition was not public knowledge.

Recall that the set of feasible divisions of lottery tickets and subjects' preferences over them were public knowledge, so that it is natural to assume complete information in modeling the bargaining game.

The nonnegligible frequency of disagreements is then incompatible with explanations based on Nash's bargaining solution or the subgame-perfect equilibrium of an alternating-offers model.

But recall that there is also a continuum of inefficient mixed-strategy demand-game equilibria, with positive probabilities of disagreement.

Roth (1985) “Toward a Focal-Point Theory of Bargaining” uses these equilibria to propose a simple miscoordination explanation of the patterns of disagreements.

His explanation focuses on mixed-strategy equilibria in which players’ beliefs are focused on the norms subjects’ public knowledge allowed them to implement, in effect reducing the demand game to a 2x2 game.

These equilibria yield disagreement frequencies close to those observed across treatments in the experiments.

However, a subsequent, more comprehensive experiment showed that this model does not fully explain how disagreement frequencies vary with the environment (Roth et al. 1988; Roth 1995b, pp. 309-311).