ABC on Deals*

Martin Dufwenberg
University of Arizona
University of Gothenburg

Maroš Servátka
University of Canterbury

Radovan Vadovič
ITAM - CIE

October 10, 2011

Abstract

We develop, and experimentally test, a behavioral model of deal-making which includes binding contracts and informal agreements as distinct but related special cases. The key behavioral assumptions are: people are mostly honest; they suffer costs of overcoming temptation to renege; and they like to split gains down the middle.

*We would like to thank Jim Andreoni, Uri Gneezy, Joel Sobel, Johan Stennek, and several participants in seminars for helpful comments and discussion. Martin Dufwenberg thanks the NSF for financial support. Radovan Vadovič thanks the Asociación Méxicana de Cultura for financial support.

1 Introduction

Bargaining theory, e.g. building on Nash (1950, 1953), focuses on binding contracts. Less attention has been given to informal (non-binding) agreements. A likely reason is that if people maximize own income, a common assumption, then there is limited scope for informal agreements to have impact.[1] A selfish agent would simply renege if this were in her interest.

Humans have tendencies that curb such opportunism. Successful entrepreneur and business leader Karl Eller, for example, wrote a book called *Integrity Is All You've Got* (2005) in which that message is clear. To some extent one can justify a preference for honesty with reference to repetition or reputation, but that cannot be the whole story. Eller writes about "the happiness that comes with knowing you'll never be ashamed to face yourself in the mirror" (p. 103). Indeed, experiments indicate that honesty matters even in settings where repeated game considerations are absent and anonymity guaranteed. Most to the point, in regards to informal agreements, Kessler & Leider (2011) find that subjects who were offered an opportunity to enter an exogenously given informal agreement did so and many subjects honored that agreement although they could have profitably reneged.[2] This intriguing finding raises the issue of how informal agreements are influenced by the economic environments in which they form.

We develop and experimentally test a model of deal-making which encompasses both informal agreements and binding contracts as distinct but closely related special cases. We assume that most people are honest, so informal agreements are mostly honored just like binding ones. From the viewpoint of what economists traditionally assume, this assumption may seem extreme but thinking about many folks around us we actually feel it makes a lot of

[1]Informal agreements may help players achieve Pareto improvements in games with multiple Nash equilibria; see e.g. MacLeod & Malcolmson (1989) and McCutcheon (1997).

[2]Relatedly, several studies have shown that people may have preferences to not deceive or to honor promises. See for example, Kerr & Kauffman-Gilliland (1994), Ellingsen & Johannesson (2004), Gneezy (2005), Charness & Dufwenberg (2006, 2010), Vanberg (2008), Deck, Servátka & Tucker (2011) and Servátka, Tucker & Vadovič (2011).

sense. Be that as it may, the assumption is testable and our experiment will shed light on to what degree it is supported.

Assuming that honesty is prevalent does not imply that informal agreements look the same as binding contracts. In this paper we focus on two issues that may affect the formation of informal agreements but are irrelevant for binding contracts.

First, we highlight cognitive costs of (over-coming) temptation. Lots of recent research has indicated that humans suffer costs when fighting temptations.[3] An honest person will honor the deals he strikes but he may not be immune to the costs he incurs when overcoming the temptation to renege. The anticipation of such costs may affect his evaluation of what deals are worth striking. Second, while we assume that most people are honest, we do not assume that all are (besides all the honest folks we know, we also met, or read about, some confidence tricksters too). When striking an informal agreement, a party has to figure in the costs associated with opportunistic counter-part behavior, which again may affect his evaluation of what deals are worth striking.

Apart from responding to these two considerations – the costs of temptation and opportunism – we propose that informal agreements look like binding contracts. So how do binding contracts look? Of course, classical bargaining theory provides some answers. However, we are skeptical of the relevance of this literature. Much of it assumes that players know each others' utilities. As an empirical proposition this is questionable. This was prominently recognized early on – *e.g.* Luce & Raiffa (1957, p. 134) find it "extremely doubtful" – but over time scholars seem to have not worried too much. Again this may be because applied economists often accept the idea that utility equals (or is monotone in) own money, in which case the issue is largely moot. However, for honest deal-makers own monetary gain is not all that matters. More generally, modern research on social preferences shows

[3] See e.g. Fudenberg & Levine (2006) and, for a survey of theoretical literature, Lipman & Pesendorfer (2011).

that people often care not only about their own monetary rewards. Experimental research has revealed some systematic insights regarding what people care about instead; nevertheless there is a lot of debate and little consensus on these matters.[4]

If human motivation is complicated, rich, and hard-to-pin-down in mathematical formulae, perhaps one shouldn't be surprised if humans strike deals that refer to readily observable data like dollars earned, rather than to some elaborate utilities? Indeed, many empirical studies suggest that agreements often simply divide dollar gains equally. Andreoni & Bernheim (2010, p. 1607) cite many studies that report such findings regarding joint ventures between corporations, share tenancy in agriculture, bequests to children, negotiation and arbitration, business partners splitting earnings from joint projects, and friends splitting tabs. Some lab experiments point in the same direction. Binmore, Shaked & Sutton (1989) run an experiment where subjects bargain over sums of money. If one of the parties breaks the negotiations both get certain outside option payoffs. It turns out that these outside options do not influence the agreements, except as imposing lower bounds on what each party will get. The money is simply split equally, as long as an equal split gives each party at least his outside option. Binmore *et al* call this a "deal-me-out solution."[5]

We assume that binding contracts generate equal splits of dollar gains, in a way akin to the deal-me-out solution.[6] We then propose that informal agreements work the same way except that they take into account costs of temptation and uncertainty related to opportunism.

The framework in which we formulate these statements precisely is game-

[4]For surveys, see Camerer (2003), Fehr & Schmidt (2006), and Cooper & Kagel (2009).

[5]Binmore et al's evidence may be read as counter to how "threat points" are typically predicted to affect bargaining outcomes according to classical theory (e.g. Nash 1950, 1953; Kalai & Smorodinsky 1975; Kalai 1977).

[6]What this means is that gains are measured relative to what the parties would have gotten had they never met rather than relative to what would happen if they did not agree once they now have met. We postpone an elaboration on the nuances involved here, as applied to real world as well as lab settings, in more detail in section 2.

theoretic and structured as follows:

1. An agreement, be it binding or informal, may be reached by a group of players about to play a game.

2. The object of negotiation concerns which strategy profile to play.

3. If the agreement is binding (a contract) then the players have to play accordingly.

4. If the agreement is informal then the players still have to play the game.

5. If no agreement is reached then the players still have to play the game.

A key feature here is the explicit anchoring on an "underlying game" (form) that has to be played. We interpret this game as describing the strategic environment that involves some business opportunity. The underlying game allows us to be explicit about the nature of the economic situation in which agreements may be struck, and it allows informal agreements to be truly non-binding (since post-agreement the game must still be played). The approach is, however, not idiosyncratic to informal agreements; as seen in 1-3 & 5, it allows for binding contracts as well.

The underlying game does not include a description of the players' haggling regarding which strategy profile they should agree to. Agreement-formation is instead implicitly captured through a solution-concept. Formally, if S is the set of strategy profiles in the game to be played then we select three elements $a, b, c \in S$ such that a is the *agreement*, b is the *behavior* following the agreement, and c the *conflict* outcome that would happen if the players failed to agree. We refer to the triple (a, b, c) as a *deal*. Our theoretical predictions are formulated in terms of restrictions on a, b, and c.

In envisaging agreements as strategy profiles we connect to a classic tradition in game theory. Von Neumann & Morgenstern (1944) approach all games other than two-player zero-sum ones with this outlook (see e.g. pp. 223-4). Nash (1953) assumes that players strike binding contracts regarding which strategy profile to play in some underlying game, and before negotiations start they simultaneously state "threats" formulated as strategies in the underlying game which an "umpire" forces them to implement if they subsequently fail to reach an agreement.[7] We share the outlook that strategy profiles are objects of negotiation, but we neither limit attention to binding contracts nor presuppose access to an umpire who enforces threats.

When it comes to contributions that consider informal agreements in an underlying game we know two related papers. We already mentioned Kessler & Leider's (2011) experimental study of exogenously given informal agreements. As regards theory, Miettinen (2011) asks, for a class of games with ordered strategy spaces, which informal agreements players will honor if they have "costs of breaking agreements". Our contribution differs from Kessler & Leider's and Miettinen's in that the shape of predicted informal agreements is endogenous rather than exogenous to the analysis.

We develop theory in section 2, report on our experiment in section 3, and conclude in section 4.

2 Theory

This section describes and interprets the game-theoretic framework in (2.1) and develops our solution concept (2.2). (The game we explore in the experiment will not be introduced and motivated until section 3.)

2.1 General Framework

Our analysis of a bargaining scenario starts with a two-player extensive game form Γ with given material (dollar) payoffs specified at end nodes. Let S_i player i's set of (behavioral) strategies (taken to be singleton if i owns no information set), and $S = S_1 \times S_2$. Let $m_i : S \rightarrow \mathbb{R}$ be i's dollar-payoff-from-strategy-profile function (derived from Γ). We'll write $m_i(s \backslash s'_j)$ to indicate

[7] See Kalai (1977) and Kalai & Tauman Kalai (2010) for more work in this vein.

i's dollar payoff if j unilaterally deviates from strategy profile $s = (s_1, s_2)$ to use s'_j instead of s_j.

This underlying game Γ describes the strategic structure of a situation where two or more persons, who just "met," face some opportunity of collaboration for mutual gain. The payoffs represent dollar increments relative to whatever wealth the players had before. If a player gets a payoff of 0 when the underlying game is played, we thus interpret this to mean that his overall dollar wealth remains the same as if he had never met the other player.

We assume that the underlying game allows enough opportunities for the players to transfer money between each other that they can achieve the following: Suppose that $\$_1, \$_2 \geq 0$ and that $\$_1 + \$_2$ equals the highest sum of all players' dollar payoffs at any endnode in Γ. We assume that Γ admits some endnode with dollar payoffs $(\$_1, \$_2)$. This implies, among other things, that Γ admits infinite choice sets and that there is a way to play Γ that results in equal dollar payoffs for all players.

We assume that Γ is a multi-stage game form with observed actions (see Fudenberg & Tirole 1991, Chapter 3), so that all instances of imperfect information concern simultaneous choices. This restriction simplifies the key definitions below (by allowing us to refer to subgames in a useful way) without essentially compromising the scope of the model since most applied and experimental work is concerned with such games.

The two players haggle about which strategy profile to play. However, Γ does not give an explicit description of this process. Rather, we capture its effect through a solution concept with a special structure. An agreement, which may be formal or informal, takes the form of a strategy profile which we denote $a \in S$. Of course, a describes also off-path play. The interpretation differs depending on whether a is an informal agreement or a binding contract. In the former case the off-path part of a reflects the players' agreed upon understanding (presumably obtained through the un-modeled haggling process) of what would happen following any deviation. In the latter case the off-path part of a should be neglected since the involved choices are inconceivable (think of a as representative of the equivalence class of strategy profiles that generate the same path).

When the negotiation phase is over the players have to play the game. What will they do? Our framework asks for specification of two more strategy profiles $b \in S$ and $c \in S$. Profile $b \in S$ describes what happens after agreement $a \in S$ is struck and $c \in S$ describes what would happen if negotiation failed. $b \in S$ and $c \in S$ also describe off-path play, again interpreted as reflecting the players' agreed upon understanding of what would happen following any possible deviation.[8]

Several clarifying comments are warranted. First, we theorize only about what happens when negotiations generate agreements. $c \in S$ should thus be understood counter-factually; it is a statement about what would happen had $a \in S$ not occurred, in the case that $a \in S$ does occur. As we shall see below it is possible that in some game no deal (a, b, c) exists which satisfies some set of properties deemed desirable. The interpretation would be that no agreement would be reached in that game if deals were assumed to possess those properties. We offer no explicit prediction for such a case of non-counterfactual negotiation-breakdown.

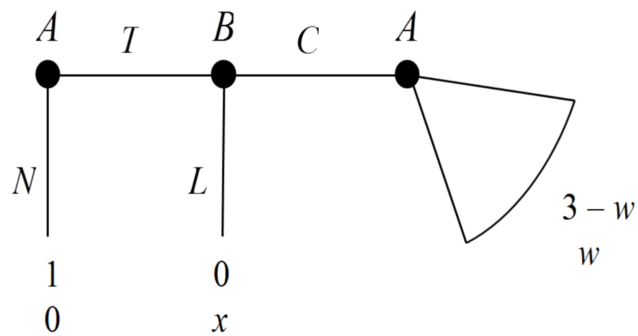
Second, with binding agreements $b = a$ by definition, but with non-binding agreements it is conceivable that $b \neq a$. Third, as regards c, one could imagine something more general, reflecting how negotiations stranded, and making c depend on such details (e.g., which player caused the break down). We do not consider such nuances. Fourth, as already indicated, we interpret off path-parts of a, b, c as reflecting common understanding by all players regarding possible off-path behavior. One could imagine alternatives, e.g., as in theory of self-confirming equilibrium (Fudenberg & Levine, 1993), but we do not do so in this paper.

Fifth, let us finally elucidate why we do not explicitly model the strategic interaction of the pre-play negotiation stage. Consider the example in Figure

[8] With binding agreements $b = a$ and off-path play is by definition not possible so (as with a) the off-path part of b should be neglected in this case.

1, where $1 < x < 3$, which comes with a story:

Figure 1: Contract game



Player A is hospital and Player B is a radiologist on payroll. A new radiography technique has been made available, and at the root Player A decides whether (T) or not (N) to invest in costly training for Player B to learn the new methods. In the former case Player B becomes more productive but also more attractive to other hospitals; choice L with subsequent payoffs reflect what happens if Player B leaves his position and takes employment at Johns Hopkins. That would be bad for Player A who stands to gain if Player B instead continues at the current job. In that case, Player A can choose what wage $w \in [0, 3]$ to pay Player B, thereby determining Player B's life-time income.[9]

The point of this example is that the game, like many others, can describe a *somewhat meaningful* situation where deals can be made. However, that situation also seems unrealistically barren, as it incorporates no opportunities for wage negotiation, promises, threats, etc. A *more meaningful* situation would arguably be the same one augmented so that the players can meet and

[9]The implicit assumption is that later in B's life he has fewer outside opportunities and is therefore vulnerable to hold-up

strike agreements regarding whether Player A should pay for the training, and what Player B's pension should be.

How should one model such considerations? How should one make predictions? One possibility may be to change the game, to include every offer, every counter-offer, every promise, every threat, every signing on a dotted line, etc., as explicit choices in a larger game. But that game is likely to get unmanageable. The player's strategy sets would be very complex. It may be intractable to apply an adequate solution concept. It is against this backdrop that we propose our approach. It is reminiscent of cooperative game theory in that we do not model every aspect of the negotiation protocol, yet it is non-cooperative in that the strategic structure of the situation to which we add a negotiation possibility is allowed to influence our analysis.

What agreements may one expect to occur? That depends on many things, including the nature of the players' motivation, whether the agreement is binding or non-binding, and game details such as x. We propose a generally applicable theory in section 2.2.

2.2 The Specific Model

What psychological and economic principles determine a deal $(a, b, c) \in S^3$? Many answers are conceivable. We propose a particular one, as follows.

$c \in S$: *Selfish SPE*

Before we describe the agreement ($a \in S$) and post-agreement behavior ($b \in S$), let us ponder what would have happened should negotiations fail ($c \in S$). We assume that players would behave selfishly in the sense that $c \in S$ is a subgame perfect equilibrium (SPE) of Γ using the dollar payoffs.

This assumption is not obvious, since it is known from a lot of studies that players are often not selfish (cf. footnote 4). However, we defend our assumption as follows: On the one hand, in many game forms (e.g. prisoners' dilemmas or trust games or public goods games) that exhibit a tension between individual and collective dollar-payoff-maximization, subjects manage to reach efficient outcomes. This suggests that players often appreciate the well-being of others. On the other hand, it doesn't seem unreasonable to suppose that players who would not manage to agree would end up being irritated with one another. All in all, it seems unclear whether as regards $c \in S$ one should assume that players would have been friendly or hostile towards one another. Our assumption that players simply would act selfishly takes a middle road.[10]

$b \approx a$: Pacta Sunt Servatka

With a binding contract, $b = a$ by definition. If an informal agreement (a, b, c) is struck, we assume that most players i simply honor the agreement and choose $b_i = a_i$. Such honesty can be justified with reference to a variety of social preferences. For example, if one thinks of informal agreements as embodying a process where players make promises to one another, then several authors have suggested that players may prefer to keep promises or (more generally) not to have lied.[11] One may alternatively imagine that players simply obey some social norm that says that one should honor agreements.[12] Or players could be guilt averse so that they are inclined not to hurt others relative to what these folks expect to get;[13] negotiation may then foster common and self-fulfilling expectations that an informal agreement will be honored by all parties. Honesty may conceivably also be backed up by reciprocity theory, if the agreed on strategy profile plus correct beliefs imply that the players are reciprocating each others' kindness.[14] So there may be multiple psycho-foundations for why agreements are honored. While it is an interesting empirical issue to determine which story may be most empirically

---

[10] Note that there is no data to guide our modeling choice: existing data on the relevance of social preferences typically do not concern games played after negotiations break down, and never concern what would happen after counter-factual negotiation breakdown.

[11] See e.g. Ellingsen & Johannesson (2004), Gneezy (2005), Demichelis & Weibull (2008), Miettinen (2008), Kartik (2009), Sutter (2009), Vanberg (2008).

[12] See Elster (1989) for a discussion of norms.

[13] See e.g. Charness & Dufwenberg (2006), Battigalli & Dufwenberg (2007, 2009).

[14] See Rabin (1993), Dufwenberg & Kirchsteiger (2004), Falk & Fischbacher (200**6**).

relevant, in this paper we do not pursue that line but rather just assume that agreements are mostly honored.

What choices should one expect from those who renege? Are they perhaps spiteful and minimize others payoffs? Do they trade off their own gain against the loss of their co-players? While these possibilities certainly do not seem unreasonable, we shall assume that renegers simply are selfish and maximize their own payoff. This way we obtain fairly easy to state definitions (below), and a clear benchmark to test with data.

$a \in S$: *Equal Gains*

Recall our remarks from the introduction regarding classical bargaining theory: While that scholarship is elegant, and conceptually related to our (a, b, c)-approach, it is of questionable applied relevance. Most classical models presume that utilities are commonly known, but moderns research on social preferences has made it clear (in our view) that utilities are nowhere near commonly known. Isn't it reasonable, then, that bargaining partners give scant reference to such complicated, rich, and hard-to-pin-down notions? Is it not likely that they instead strike deals that, most of the time, refer to readily observable data like dollars earned? This is the view we take in this paper.

To arrive at specific predictions we take inspiration from the so-called deal-me-out solution, discussed by Binmore *et al.* Recall their definitions and experimental findings: They look at negotiation between two players who may divide a sum of money. The bargaining rules include a specification of how much money each player gets if negotiations strand ("the outside option payoffs"). The deal-me-out solution says that players split the available money equally, as long as each of them gets no less than his outside option. In the experiment, by and large, Binmore *et al* report experimental support for this solution.[15]

[15] Our short account here does not give full account to Binmore *et al*'s rich and fascinating data set, but for details and caveats we refer the interested reader to the original source.

The solution we propose may be conceptually viewed as an extension of the deal-me-out solution. It modifies or adds to Binmore *et al*'s notion in two ways: First, we assume that the object of the negotiation is a strategy profile ($a \in S$) rather than a sum of money. Second, we do not restrict attention to binding contracts but also consider informal agreements.

We propose that deals generate equal splits of dollar gains, as long as each player i makes a dollar gain relative to when play proceeds according to $c \in S$. If the deal involves a binding contract then what this means is straightforward: i gains if $m_i(a) > m_i(c)$. (Think of $m_i(c)$ as the counterpart in our theory to Binmore *et al*'s outside option for i.)

If the deal involves an informal agreement then it is not as obvious whether or not i makes a dollar gain relative to $c \in S$. There are (at least) two reasons. First, there are the dishonest folks who renege. Suppose that player i believes that with probability $\varepsilon_i > 0$ his co-player j will renege and choose selfishly; call that strategy $\hat{s}_j \in \arg\max_{s_j \in S_j} m_j(a \backslash s_j)$.[16] Then i's expected dollar payoff under an informal agreement will be lower than with the same agreement under a binding contract. If the difference is big enough, i may accept an equal-split inducing strategy profile $a \in S$ as a binding contract but not as an informal agreement.

A second consideration complicating how a player may view his gain concerns the temptations to renege that he has to overcome if he honors an informal agreement. There is a sizable literature on human tendency to resist (or subdue) temptations.[17] It is often argued that humans can overcome temptation, but that this comes at a cost. If a player i considers such costs when evaluating an informal agreement, then his subjective dollar gain (i.e., net of the temptation cost) under an informal agreement will be lower than

[16] This formulation assumes that a selfish player j does not take into account the possibility that his co-player reneges too, a feature we could easily change with little conceptual gain and much added notational complexity and no import whatsoever as regards our upcoming experimental games of Section 4.

[17] See Lipman & Pesendorfer (2011) for a survey. Numerous economic experiments have demonstrated the impact of temptations on behavior. Two examples are Benartzi & Thaler (2004) and Brown, Chua & Camerer (2009).

with the same strategy profile as a binding contract. Again, if the difference is big enough, i may accept an equal-split inducing strategy profile $a \in S$ as a binding contract but not as an informal agreement.

How should one calculate costs of overcoming temptation exactly? Are they linear or perhaps convex in how much a player may gain by reneging (cf. Fudenberg & Levine 2006 who discuss both versions)? Are they stochastic (cf. Dekel & Lipman 2011)? Do they depend on how many times along a path a player is tempted, or does only the maximum temptation along the path matter? Are the cost moderated to the extent that reneging hurts others (as would seem consistent with the evidence of Gneezy 2005)? The answers are by no means obvious (although for many games, including those of section 4, they do not influence predictions). We go with the following straightforward version that allows us to state definitions fairly easily. The cost of overcoming temptation is proportional to the maximal forgone dollar gains from reneging; let the proportional factor of player i be $\gamma_i > 0$.[18] With an informal agreement, i's cost-of-temptation associated with $a \in S$ equals

$$\gamma_i \times (\arg\max_{s_i \in S_i} m_i(a \backslash s_i) - m_i(a)).$$

With a binding contract, by contrast, temptation-costs equal zero; there cannot be any reneging on a binding contract and hence no temptation-cost.

We next give definitions that formally pin down, respectively, binding contracts and informal agreements. We then offer further comments on interpretation and the predicted difference between binding contracts and informal agreements.

- A binding contract $a \in S$ satisfies:

  (i) for $i = 1, 2$ and some $M > 0$ we have

$$m_i(a) = \max\{M, m_i(c)\},$$

[18] Fudenberg & Levine's Definition 2 is similar in spirit. We adopt the notation γ_i from them.

(ii) M is the largest number for which (i) holds,

(iii) off its induced path, a prescribes SPE-play using the dollar payoffs.

- An informal agreement $a \in S$ satisfies:

  (i) for $i = 1, 2$ and some $M > 0$,

$$m_i(a) - \gamma_i[\arg\max_{s_i \in S_i} m_i(a \backslash s_i) - m_i(a)] \\ - \varepsilon_i[m_i(a) - m_i(a \backslash \hat{s}_j)] = \max\{M, m_i(c)\},$$

  where $\gamma_i, \varepsilon_i > 0$ and $\hat{s}_j \in \arg\max_{s_j \in S_j} m_j(a \backslash s_j)$,

  (ii) M is the largest number for which (i) holds,

  (iii) off its induced path, a prescribes SPE-play using the dollar payoffs.

Look at the binding contract definition first. Condition (i) embodies our take on the deal-me-out feature; players simply split dollar gains equally, unless someone would then get less than he would at c. Condition (ii) is an efficiency condition such that no money is left on the negotiation table; obviously $M = \max_{a \in S} \Sigma_{k \in N} m_k(a)$. Condition (iii) is redundant in the sense that off-path play is inconceivable with a binding contract, but needed to make $a \in S$ formally well defined.

Now look instead at the informal agreement definition. Conditions (i) and (ii) again capture the deal-me-out idea. Comparing to the binding contract case, one sees that i's gain relative to $m_i(c)$ now subtracts the costs of overcoming temptation and co-player opportunism.[19] Note that (i) approaches the corresponding condition for a binding contract, if γ_i, $\varepsilon_i \to 0$.

[19] Condition (i) reflects the idea that player i factors in temptation costs evaluated with respect to a fully honest opponent, independently of ε_i. This makes sense if one assumes that any temptation cost is borne (only) at the point of agreement, before any actions are taken. This modeling has the advantage of allowing a straightforward mathematical formulation, and it is in line with the modeling choices of Fudenberg & Levine.

In this sense, relative to informal agreements, binding contracts "switch off" the costs of temptation and opportunism. Condition (ii) is again an efficiency condition, but this time it may be that $M < \max_{a \in S} \Sigma_{k \in N} m_k(a)$ (the reader may verify this by constructing an example). Condition (iii) embodies an assumption about play after a player reneges. We suggest that the justification is analogous to what lead us to assume that $c \in S$ is a subgame perfect equilibrium using the dollar payoffs; we refer to our earlier discussion of that matter.

The informal agreement definition implicitly assumes that γ_i and ε_i are commonly known by the bargaining parties (how else could these parameters pin down the agreements so precisely?). We also like to think that γ_i and ε_i differ across players (hence the subscript i) so that the nature of a deal will vary across negotiation instances. There is an obvious tension between those two sentences! How can this be reconciled? We interpret our solution to imply that the bargaining parties simply reveal γ_i and ε_i to each other in the course of negotiations. After all, they are (most of the time) honest!

Note that the predicted agreement in many cases does not change with small changes in γ_i and ε_i. For example, if γ_i and ε_i are 'small' and there is a reasonably 'large' distance between $m_i(a)$ and $m_i(c)$, then the the players will simply go for the equal split in all cases.

$a, b, c \in S$: Square Dealing

It is time to sweep everything together in our key definition. According to dictionaries, "square" can mean "straightforward and honest." It can also mean (in math) that all sides are equal. Since we assume that most folks are honest and that many deals involve straightforward equal splits, we find it appropriate to call our predictions square deals:

Definition: The triple $(a, b, c) \in S^3$ is a *square deal* if

(A) a satisfies the parts (i), (ii), (iii) of whatever case (informal agreement or binding contract) is relevant,

(B) If $i = 1, 2$ is honest, then $b_i = a_i$. If $i = 1, 2$ is dishonest then he chooses $\hat{s}_i \in \arg\max_{s_i \in S_i} m_i(a \backslash s_i)$,

(C) c is an SPE of Γ using the dollar payoffs.

In section 3 we will apply the Definition to a particular game and test the predictions of our theory in an experiment.

3 An Experiment

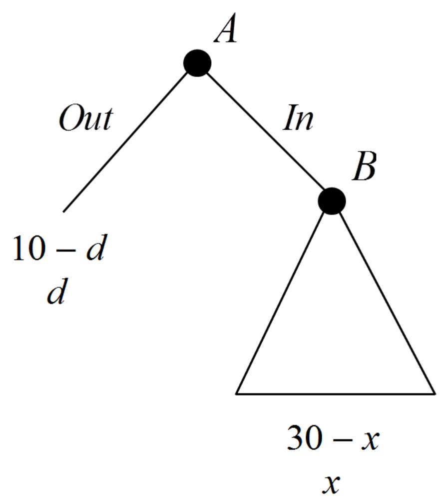
This section reports on an experimental test of the model in section 2.2. We first discuss the games on which our design is based and derive predictions (3.1), and then the experimental design and procedures (3.2) and results (3.3).

3.1 Games and Predictions

The model of section 2.2 generates qualitatively different predictions for binding contracts and informal agreements. The difference can be traced to costs of temptations and beliefs about the likelihood that the co-player reneges, issues that come up only when agreements are informal.

We use the lost wallet game (Dufwenberg & Gneezy, 2000) presented in Figure 2, which may be viewed as a simplified version of the doctor-hospital game that was presented in section 2.2. It is appropriate for our design in the sense that it possesses several key qualities: 1. The game is simple to explain to subjects and implement in a laboratory setting, yet rich enough to allow a deal with equal split. 2. The theory is simple to apply and generates sharp comparative statics predictions for binding contracts versus informal agreements as well as for a payoff treatment variation. 2. 3. It has a unique SPE using the dollar payoffs, and so a unique square deal. 4. Players only have to consider ε_A and γ_B (and can ignore ε_B and γ_A), which substantially reduces complexity regarding unobservables.

Figure 2: The lost wallet game



The game is played by two players, A and B. Player A chooses between In and Out. If he chooses Out, Player A and Player B receive their respective payoffs $10-d$ and d (where $d \in \{0, 5\}$ is a parameter). Alternatively, if Player A chooses In then Player B divides \$30 between the two players. Player B can keep any amount $x \in [0, 30]$ and transfer amount $30-x$ to Player A. The game has a unique subgame perfect equilibrium $(Out, x = 30)$ with payoffs $(5, 5)$ and $(10, 0)$ depending on d. In the rest of this section we adopt the following convention: we will refer to y as the agreed-upon x and to z as the post-agreement choice of x.

Suppose that before playing this game the parties could negotiate over what each of them should do and possibly arrive at an agreement. The agreement can either be binding or informal and our theory of square deals gives predictions for each case. The disagreement payoffs are the SPE payoffs: $m_A(c) = 10 - d$ and $m_B(c) = d$. Condition (A) of the Definition implies that the agreement will take the form (In, y), where y must satisfy conditions (i)-(iii) of either binding contract (BC) or informal agreement (IA) case.[20]

Binding Contracts: A binding contract rules out reneging by definition. This immediately takes temptations as well as any profiteering behavior by opportunistic types out of the picture. The theory predicts that players will agree on an equal split of 15 for each player (since $15 > m_i(c)$ for $i = A, B$; cf. condition (i)): $m_A(a) = m_B(a) = 15$.

Informal Agreements: For sufficiently low ε_i's and γ_i's we obtain the same predictions as in the BC case – namely $m_A(a) = m_B(a) = 15$.[21] Otherwise, if either ε_i's or/and γ_i's are large, we get either compensated deals or no deals at all. Part (i) of the IA case implies that Player A will demand compensation (an amount greater than 15) when

$$\begin{aligned} 15(1-\varepsilon_A) &< 10-d \\ \varepsilon_A &> (5+d)/15. \end{aligned}$$

The compensated deal allocates Player A an amount

$$m_A(a) = \frac{10-d}{1-\varepsilon_A}.$$

Notice however, that there is a limit on how much Player A can demand from Player B. Because compensation is a direct transfer from Player B to Player A a too high demand may not be feasible in the sense that Player B would eventually reject it, i.e., conditions (i) and (ii) would fail to hold. Thus, Player B is willing to compensate Player A if

$$\begin{aligned} (30-\frac{10-d}{1-\varepsilon_A}) - \gamma_B(\frac{10-d}{1-\varepsilon_A}) &> d \\ \gamma_B &< \frac{20-\varepsilon_A(30-d)}{10-d}. \end{aligned}$$

[20] In one of our treatments IA[5,5] the theory allows a degenerate deal in which $a = c$. In what follows we will ignore this type of deal (it also gets no support in the data) and focus on efficient deals involving Player A choosing In.

[21] Recall from our previous discussion that as ε_i's and γ_i's all go to zero the IA case converges to the BC case.

Now we turn to Player B. Player B will honor any agreement and keep only the agreed-on amount y. But, as we discussed, honoring agreement comes at the cost of overcoming temptation. Player B may need to be compensated if his temptation is sufficiently high,

$$\begin{aligned} 15-\gamma_B 15 &< d \\ \gamma_B &> (15-d)/15, \end{aligned}$$

in which case he will demand

$$d+\gamma_B 15.$$

But Player B also cannot demand too much – otherwise, condition (i) and (ii) may violated. The deal will be accepted by A only if

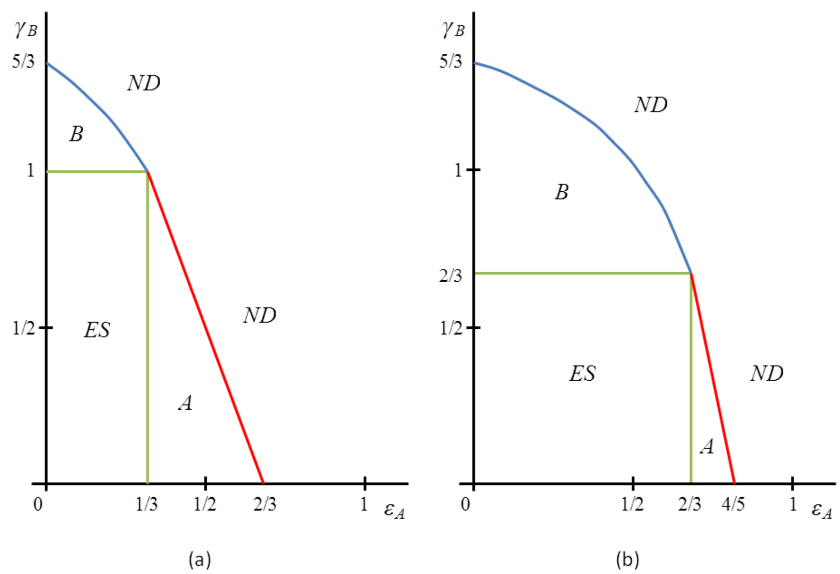
$$\begin{aligned} (30-(d+\gamma_B 15))(1-\varepsilon_A) &> 10-d \\ \gamma_B &< \frac{20+\varepsilon_A(30-d)}{15(1-\varepsilon_A)}. \end{aligned}$$

The predictions are summarized in the Figure 3 below. ES is the region in which deal amounts to an equal split; A is the region in which Player A gets more than 15 (is compensated); B is the region where B gets more than 15; and ND is region with no deal.

Panel (a) refers to the case when $d=0$ and panel (b) to the case when $d=5$. Notice how the regions with equal splits (ES) and compensated deals (A, B) change as the disagreement payoff varies between $(10,0)$ to $(5,5)$. When $d=0$, Player A can get 10 by choosing Out. An agreement on equal split would give him 15 but one has to remember that that 15 is not guaranteed. It only happens if Plater B does not renege, i.e., with probability $(1-\varepsilon_A)$. If ε_A is high the expected payoff from an equal split agreement falls short of the sure 10 in which case Player A needs to be compensated. Thus Player A's high disagreement payoff expands the region in which he needs to be compensated (region A).

On the other hand, as the disagreement payoff becomes $(5,5)$, two separate effects take place. First, Player A now accepts an equal split agreement

Figure 3: Predictions



even for higher ε_A because his secure payoff from Out is not as high; and second, since Player B can now insure at least 5 by dropping the negotiations, and equal split net of his temptation cost may not be good enough without an additional compensation. This is reflected in panel (b) by region A being smaller and region B being larger relative to panel (a).

3.2 Design and Procedures

The computerized experiment was conducted at the University of Arizona in the Economic Science Laboratory. The software was written in Visual Basic 6. In total, 204 undergraduate students participated in the experiment as subjects (the sessions and participation is summarized in Table 5). Once all subjects were seated at computer terminals separated by privacy dividers, they were handed out hard copies of instructions (see the Appendix) and given 10-15 minutes to read them on their own. When everyone was finished reading, the instructions were read aloud for the whole group. After this, the experimenters individually answered any questions. The software started up with a set of comprehension questions (see the Appendix). Every subject had to answer all questions correctly for the experiment to proceed further. On average, a session lasted about 50 minutes. The average final payment was \$19.90, including a \$5 show-up fee.

The theory presumes pre-play negotiation but leaves this process implicit, reflected only through the solution concept. In the lab we have to offer some specific format for the haggling, however, and we chose an alternating-offer structure. After being acquainted with the details of the game and learning their respective roles as Player A or Player B, the subjects entered a negotiation stage in which they could send proposals back and forth and agree on how to play the game. One person from each pair was randomly selected to make an opening proposal. Each proposal specified whether Player A would choose In or Out, and, conditional on In, the amount, y, that Player B would keep. The party who received a proposal could (i) accept it; (ii), make a counter-proposal, or (iii) disagree and quit negotiating. Acceptance of a proposal led to an agreement. This ended the negotiations and caused a message saying either "Player A chooses OUT" or, for $x = 18$, "Player A chooses IN and Player B keeps \$18 and gives \$12 to Player A" to appear on the pair's computer screens. A counter-proposal reversed the negotiation roles while a disagreement terminated the negotiation process. Following the negotiations subjects played the game in a manner that differed between treatments.

To summarize, our experiment uses a 2×2 design. We vary the type of agreement, whether it is a binding contract (BC) or an informal agreement (IA), and the payoff following option Out to be either $(10, 0)$ or $(5, 5)$. If no agreement was made the paired subjects entered the game stage where they were free to make any decisions. With informal agreements (IA treatment) agin no restrictions were imposed on the game-choices. With binding contracts (BC treatment), the agreements from the negotiation stage were automatically implemented.

3.3 Results

The raw data contrasting agreements and actual behavior are presented in Table 1. This provides a nice visual overview and allows one to get the full flavor of the data. The theory of section 2.2 generates clear predictions for our design, outlined in section 3.1. The first prediction is that agreements form in the BC treatments. In the IA treatments agreements only form if subjects' temptations (γ_i's) and estimates of others' honesty (ε_i's) are sufficiently low.

Let us look at agreements more closely. Table 2 shows that apart from two cases in IA[5,5] all other pairs of subjects reached an agreement. In addition, all but one agreements involved Player A choosing In.[22] This is very much expected in the BC treatments but in the IA treatments temptations or low belief in honesty could have easily led subjects to an impasse or to dropping the negotiations altogether. The fact that 100% of our subject-pairs agreed in both of the IA treatments implies that, if the theory is valid, honest subjects' γ_i's and ε_i's were reasonably low.

Result 1: Subjects overwhelmingly arrive at agreements: 100% in IA and 96% in the BC treatments. Agreements are efficient (99% involve Player A choosing In).

The theory presumes that most players are honest and Table 2 shows that a majority of agreements were indeed honored. All Player As stayed true to their word and chose In. This implies that subjects themselves must have had a high belief in Player B's honesty (i.e., low ε_i). Player Bs faced a dilemma. In order to stick with their agreements they had to walk away

[22] One pair agreed on *Out* but perhaps this was a mistake because then Player A chose *In* and Player B kept 15.

Table 1: Raw data on agreements and behavior

IA[10,0]		IA[5,5]		BC[10,0]	BC[5,5]
Agreement	Behavior	Agreement	Behavior	Agr=Beh	Agr=Beh
0	15	15	15	14	14
10	20	15	15	15	15
13	18	15	15	15	15
15	15	15	15	15	15
15	15	15	15	15	15
15	15	15	15	15	15
15	15	15	15	15	15
15	15	15	15	15	15
15	15	15	15	15	15
15	15	15	15	15	15
15	15	15	15	15	15
15	15	15	16	15	15
15	15	15	17	15	15
15	15	15	25	15	15
15	15	15	20	15	15
15	15	15	30	15	15
15	15	15	30	15	15
15	15	15	30	15	15
15	15	16	16	15	15
15	15	16	20	15	15
15	20	17	15	15	15
15	30	18	18	15	16
15	30	18	18	15	17
15	Out	20	20	20	20
-		20	20	-	20
-		24	22	-	Disagr.
-		Out	15	-	Disagr.

from sizable gains that they could have easily pocketed.[23] So did Player

[23]In IA[10,0] Player Bs who lived up to their agreements earned on average \$15 but could have earned \$30. They left as much as \$15 on the table. In treatment IA[5,5] Player Bs earned on average \$16.55 which means they left \$14.35 on the table.

Table 2: Agreements and honesty

Treat.	Obs	Agr	Pl. A: *In*		Pl. B: y		
			Agr	Dec	$> z$	$= z$	$< z$
IA[10,0]	24	24	24	23	0	17	6
IA[5,5]	27	27	26	27	2	16	7
BC[10,0]	24	24	24			24	
BC[5,5]	27	25	25			25	

Note: In IA[10,0] one pair has agreed on Player 1 choosing Out. Following this Player 1 chose In and Player 2 kept 15.

Bs measure up? Indeed. The proportion of honest Player Bs is quite high. In IA[10,0] treatment 74% of Player Bs did exactly as they agreed and in IA[5,5] the proportion was slightly lower at 64%. This number confirms that subjects' high belief in honesty is in fact backed up by the actual behavior.

Result 2: The majority of our subjects act as if honest: 98% of Player As and 69% of Player Bs in IA treatments behave exactly as they agreed.

Next we explore the sensitivity of agreements to variation in outside options in the binding versus informal setting. Our theory (in section 3.1) predicts that we should observe equal splits and no sensitivity to outside option in the BC treatments. In the IA treatments, on the other hand, in addition to equal splits we should also observe compensated deals. These ought to be sensitive to the treatment variation in the outside option. In the IA[10,0] we expect more compensated deals favoring Player A and less compensated deals favoring Player B than in the IA[5,5].

A quick look at Table 3 tells the picture. The distributions of agreements in BC[10,0] and BC[5,5] (the first two rows) are almost identical. In both treatments a vast majority of subjects agreed to split $30 right down the middle. Varying the payoffs following choice *Out* from $(10, 0)$ to $(5, 5)$ does not seem to have any impact on the outcome of negotiations (two-tailed Fisher's exact test $p-$value $= 0.58$). The agreements struck in the IA treatments

Table 3: Agreements

Treat.	$y < 15$	$y = 15$	$y > 15$
BC[10,0]	1	22	1
BC[5,5]	1	20	4
IA[10,0]	3	21	0
IA[5,5]	0	18	8

presented in the right panel look different. The majority of subjects still agree on equal splits: 87.5% in the IA[10,0] and 69.2% in the IA[5,5]. But, interestingly, in IA[10,0] all pairs agree on giving Player B *at least* \$15 while in treatment IA[5,5] all pairs agree on giving Player B *at most* \$15. The equality of distributions is rejected at the 1% level $p-$value $= 0.001$. Both of these observations are very much in line with the theory.

Result 3: Data support the theory of square deals. In all treatments the majority of agreements concentrate on equal splits: 86% in the BC treatments and 78% in the IA treatments. Binding contracts are not sensitive to variation in the payoff from outside option. Informal agreements show sensitivity in the predicted direction: Player As (Bs) get compensated more (less) often when their outside option payoff increases (decreases).

Although on balance our theory fares well with the data it is not spot on (as abstracting theories never are). We will now discuss two interesting features of the data that are not captured by our theory.

The first concerns reneging behavior. Recall that we assumed that most players are honest and those that are not honest are fully opportunistic. If this were the correct description of the world, then every time a subject breaks an agreement he effectively reveals he is opportunistic. And, as that, he should act in his best self-interest. That is, we should observe that any Player B who reneges also keeps all \$30 and gives nothing to Player A. About 31% (16/51) of Player Bs reneged but only five of them kept the whole amount \$30. Thus we only find five opportunistic Player Bs. The rest who renege usually give their matched Player A at least something.

The source of this behavior might lie in social preferences (Ochs & Roth 1989, Binmore 1991). However, there is also another possibility. Let us focus the attention on IA[10,0] treatment. There, we observe three compensated deals that give Player A more than 15: (0,15), (10,20), (13,18).[24] Following each of those deals Player B reneges but does not keep everything, rather just "shades" the agreed-on amount by some fraction. It is "as if" the Player B had a different deal in mind - one that compensated him. Once he has the power to decide, he goes on to implement it. In the IA[5,5] treatment only three out of eight Player Bs who negotiated compensated deals reneged. Out of those three one shaded the agreed-on amount (15,20) but the other two actually gave Player A more money: (17,15), (24,22). This just shows that most of our Player Bs who renege do not act opportunistically but rather seem to act on their own view of what constitutes a just deal.

Second, we look more closely at deals that deviate from equal split in BC[5,5]. There, one of the deals gave Player A more (\$16) but as many as four deals gave Player B more than \$15. Was this a mistake or did Player B have some legitimate claim to get the upper hand? Our theory is very clear on the issue. In the BC treatments agreements should split the \$30 right down the middle. The only legitimate claim for a compensation can be made in the IA treatments and it would imply prohibitively high temptation cost or very low belief in others' honesty. If in BC[5,5] Player B has no good reason to request more than \$15 we would expect that such claim would be rejected by the paired player and could possibly lead into a negotiation conflict. On the other hand, in the IA[5,5] a legitimate compensation claim by Player B should be relatively easy to accept.

In Table 4 we list the sequences of offer exchange for all deals that gave Player B more money in the two [5,5] treatments. A quick visual inspection

[24] The first number in the parenthesis is the agreed-on and the second the chosen amount that Player B kept.

Table 4: Sequences of offers

	IA[10,0]								BC[10,0]			
Obs:	1	2	3	4	5	6	7	8	1	2	3	4
Op. Ofr:	*B*	*B*	*A*	*A*	*B*	*B*	*A*	*A*	*B*	*B*	*B*	*B*
Rnd: 1	16	20	17	18	16	20	20	24	20	20	20	20
2		15			18				15	15	15	
3		18							17	18	20	
4		16							16	16	10	
5										18	20	
6										17	15	
7										18	20	
8										15	15	
9										18	20	
10										16	15	
11										18	20	
12										17	15	
13										18	20	
14										*Out*		
15										17		
Agr.:	16	16	17	18	18	20	20	24	16	17	20	20

confirms what we suspected. In the BC[5,5] compensated deals involved a fight and a struggle between the subjects. In contrast the deals in IA treatments were formed smoothly. This observation corroborates our Result 3 and makes us more confident that the departures from equal splits in the BC treatments are typically not found to be justified by both parties and constitute hard bargains. The resulting deal is then likely driven by an imbalance in subjects' patience and abstinence.

4 Discussion

We have proposed a theory of deal-making which covers informal agreements and binding contracts as special cases. We suggest that people – who deep down may be motivated in complicated, rich, and hard-to-pin-down-in-mathematical-formulae ways – do not engage in any elaborate utility calculus when striking deals. Rather they tend to focus on easy to observe data, like dollars gained. Inspired by Binmore *et al*'s so-called deal-me-out solution, we assume that people agree to split all dollar gains made available by their interaction equally, as long as each person is thereby made better off than if the negotiations stranded. A key assumption in our approach is that most people are honest. Informal agreements therefore work in similar ways as binding contracts. However, the two are not interchangeable. With informal agreements players incur costs of over-coming temptation and of co-player opportunism. This influences players' perception of whether a deal makes them better off than if negotiations stranded. The concern is irrelevant with binding contracts. Therefore our theory generates systematic testable predicted differences between the shapes and impact of informal agreements and binding contracts. We ran an experiment to evaluate the empirical relevance of these predictions. Modulo some caveats, we found considerable support.

Informal agreements have been given rather scant attention in previous work. Are they economically unimportant? Couldn't agents simply always rely on binding contracts to achieve good partnership outcomes? We do not think so, and there are several reasons.

First, binding contracts may be inconceivable. Think, for example, of two fishermen who live in a developing country where neither courts nor policemen are reliable. It may be impossible for these fellows to draw up a binding contract which regulates their access to a nearby little lake. Does this mean that they are destined to inefficient excessive depletion of the fish stock? The answer may be no, if they can rely on an informal agreement.

Second, binding contracts may be illegal. Think, for example, of cartels in industrial countries. Courts exist and police can be relied on, yet a price fixing agreement would not be legal. A binding contract is thus ruled out. Does that imply that market outcomes will conform with standard IO theory, where firms maximize profit taking each others' actions as given? (Think of a one-shot Bertrand model to get a concrete example.) Perhaps not. Suppose two firm representatives meet in a bar, have a pint, and shake hands and agree to choose monopoly prices. Will this stick? If people are honest, the answer may be yes.

Third, even if binding contracts are feasible in principle (as they perhaps usually are) they may be costly. Think, for example, of a guy who meets a girl and they play the (one-shot, sequential) game of life with decisions on nuptials, kids, divorce, alimony, etc. Detailing a binding contract may involve significant costs for reasons ranging from lawyers' fees to unforeseen contingencies to awkward feelings regarding legal chit-chat during courtship. Perhaps, instead, the couple will shun the formalities, look one another in the eye, and promise to be faithful forever?

It is natural to wonder just how compelling these examples are. Under which circumstances will informal agreements work? How exactly will the terms be structured? How will the deals shape subsequent play? To answer such questions one needs theory. We hope that ours will be helpful in this regard, to the extent that it is substantively generates good predictions as well as because it forms a starting point for further work in this area.

References

[1] Andreoni, J. and B.D. Bernheim, (2009): "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica* 77, 1607–1636.

[2] Battigalli, P. and M. Dufwenberg, (2007): "Guilt in games," *American Economic Review Papers and Proceedings* 97, 170–176.

[3] Battigalli, P. and M. Dufwenberg, (2009): "Dynamic Psychological Games," *Journal of Economic Theory* 144, 1–35

[4] Binmore, K., Shaked, A. and J. Sutton, (1989): "An outside option experiment," *Quarterly Journal of Economics* 104, 753–770.

[5] Benartzi, S., and R. Thaler, (2004): "Save More Tomorrow: Using Behavioral Economics to Increase Employee Savings," *Journal of Political Economy*, 112.1(2), 164-187.

[6] Bolton G. E., (1991): "A Comparative Model of Bargaining: Theory and Evidence," *American Economic Review*, 81(5), 1096-1136.

[7] Brown, A. L., Z. E. Chua, and C. F. Camerer, (2009): "Learning and Visceral Temptation in Dynamic Savings Experiments," *Quarterly Journal of Economics*, 124(1): 197-231.

[8] Camerer, C., (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton University Press.

[9] Charness, G. and M. Dufwenberg, (2006): "Promises and partnership," *Econometrica* 74, 1579–1601

[10] Charness, G. and M. Dufwenberg, (2010): "Bare Promises: An Experiment," *Economics Letters* 107, 281–283.

[11] Deck, C., (2010): "An Experimental Investigation of Trust and Sequential Trade," *Southern Economic Journal* 76, 993–1004.

[12] Deck, C., M. Servátka, and S. Tucker, (2011): "Comment on Promises and Partnership," working paper.

[13] Dekel and Lipman, (2011): "Costly Self Control and Random Self Indulgence," working paper.

[14] Demichelis, S. and J. Weibull, (2008): "Language, meaning and games - A model of communication, coordination and evolution," *American Economic Review* 98, 1292–1311.

[15] Dufwenberg, M. and U. Gneezy, (2000): "Measuring Beliefs in an Experimental Lost Wallet Game," *Games and Economic Behavior* 30, 163–182.

[16] Dufwenberg, M. and G. Kirchsteiger, (2004): "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, 47, 268–298.

[17] Eller, K., (2004): *Integrity Is All You've Got*, McGraw-Hill Companies.

[18] Ellingsen, T. and M. Johannesson, (2004): "Promises, Threats and Fairness," *Economic Journal* 114, 397–420.

[19] Elster, J., (1989): "Social Norms and Economic Theory," *Journal of Economic Perspectives* 3, 99–117.

[20] Falk, A. and U. Fischbacher, (2006): "A Theory of Reciprocity," *Games and Economic Behavior* 54, 293–315.

[21] Fehr, E. and K. Schmidt, (2003): "Theories of Fairness and Reciprocity – Evidence and Economic Applications,".In: M. Dewatripont, L. Hansen and S. Turnovsky (Eds.), *Advances in Economics and Econometrics –* 8th World Congress, Econometric Society Monographs, Cambridge University Press 2003.

[22] Fudenberg, D. and D.K. Levine, (1993): "Self confirming equilibrium," *Econometrica*, 61(3), 523-545.

[23] Fudenberg, D. and D.K. Levine, (2006): "A Dual-Self Model of Impulse Control," *American Economic Review* 96, 1449–1476.

[24] Fudenberg, D. and J. Tirole, (1991): "Game Theory," The MIT Press.

[25] Geanakoplos, J., D. Pearce, and E. Stacchetti, (1989): "Psychological Games and Sequential Rationality," *Games and Economic Behavior* 1, 60–79.

[26] Gneezy, U., (2005): "Deception: The Role of Consequences," *American Economic Review* 95, 384–394.

[27] Kalai, E., (1977): "Proportional solutions to bargaining situations: Inter-personal utility comparisons," *Econometrica* 45, 1623–1630.

[28] Kalai, A., and E. Kalai, (2010): "A Cooperative Value for Bayesian Games," Working paper 1512, Northwestern University.

[29] Kalai, E., and M. Smorodinsky, (1975): "Other solutions to Nash's bargaining problem," *Econometrica* 43, 513–518.

[30] Kartik, N., (2009): "Strategic Communication with Lying Costs," *Review of Economic Studies* 76, 1359–1395.

[31] Kerr, N.L. and C.M. Kauffman-Gilliland, (1994): "Communication, Commitment, and Cooperation in Social Dilemmas," *Journal of Personality and Social Psychology* 66, 513–529.

[32] Kessler, J.B. and S. Leider, forthcoming: "Norms and Contracting," *Management Science*.

[33] Lipman, B.L. and W. Pesendorfer, (2011): "Temptation," working paper.

[34] MacLeod, W.B. and J. Malcolmson, (1989): "Implicit Contracts, Incentive Compatibility and Involuntary Unemployment," *Econometrica* 56, 447–480.

[35] McCutcheon, B., (1997): "Do Meetings in Smoke-Filled Rooms Facilitate Collusion?" *Journal of Political Economy* 105, 330–350.

[36] Miettinen T., (2008): "Contracts and Promises - An Approach to Preplay Agreements," working paper.

[37] Nash, J., (1950): "The Bargaining Problem," *Econometrica* 18, 155–162.

[38] Nash, J., (1953): "Two-person Cooperative Games," *Econometrica* 21, 128–140.

[39] Ochs, J., and A. E. Roth, (1989): "An Experimental Study of Sequential Bargaining," *American Economic Review*, 79(3), 355-84.

[40] Rabin, M., (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review* 83, 1281–1302.

[41] Servátka, M., S. Tucker, and R.Vadovič, (2011): "Words Speak Louder Than Money," *Journal of Economic Psychology* 32(5), 700–709.

[42] Sutter, M., (2009): "Deception through telling the truth?! Experimental evidence from individuals and teams," *Economic Journal* 119, 47–60.

[43] Vanberg, C., (2008): "Why do people keep their promises? An experimental test of two explanations," *Econometrica* 76, 1467–1480.

Appendix

A Additional Tables

Table 5: Treatments

Tr. Name	Type of Agreement	Outside Opt.	Session	# of subj.
IA[10,0]	Informal	[10,0]	Sess. 1	20
IA[10,0]	Informal	[10,0]	Sess. 2	28
IA[5,5]	Informal	[5,5]	Sess. 1	28
IA[5,5]	Informal	[5,5]	Sess. 2	26
BC[10,0]	Binding	[10,0]	Sess. 1 & 2	10
BC[10,0]	Binding	[10,0]	Sess. 3	28
BC[5,5]	Binding	[5,5]	Sess. 1	30
BC[5,5]	Binding	[5,5]	Sess. 2	24

B Instructions

In what follows we present the universal version of the instructions in which {... *or* ...} always contains two different versions of the text that was used appropriately in different treatments.

Now that the experiment has begun, we ask that you do not talk with each other for the duration of the experiment. If you have a question after we finish reading the instructions, please raise your hand and the experimenter will approach you and answer your question in private.

You will receive $5 for participating in this experiment. You may also receive additional money, depending on the choices made (as described below). Your earnings will be paid to you in cash individually and privately.

During the session, you will be paired with another person. However, no participant will ever know the identity of the person he or she is paired with.

Table 6: Descriptive statistics

Tr.	Obs.	Agreement			Decisions	
		Agreed (%)	Pl. A: In (%)	Pl. B: y (st. dev.)	Pl. A: In (%)	Pl. B: y (st. dev.)
IA[10,0]	24	24	24	14.08	23	16.13
		(100)	(100)	(3.19)	(96)	(5.65)
IA[5,5]	26	26	25	16.11	26	18.54
		(100)	(96)	(2.321)	(100)	(4.99)
BC[10,0]	24	24	24	15.17	24	15.17
		(100)	(100)	(1.05)	(100)	(1.05)
BC[5,5]	27	25	25	15.48	25	15.48
		(93)	(100)	(1.45)	(100)	(1.45)

Note: In IA[10,0] one pair has agreed that Player A chooses OUT. Following this Player A chose IN and Player B kept 15. In only two instances, both in BC[5,5], subjects have disagreed. In both cases Player A chose IN. Player B kept 20 and 30 respectively.

Table 7: Frequencies at prediced values

Obs. y :	10	11 or 12	13 or 14	15	16 or 17	20
IA[10,0]	0	0	0	21	1	1
	(0)	(0)	(0)	(87.5)	(4.2)	(4.2)
IA[5,5]	2	2	1	18	0	0
	(7.4)	(7.4)	(3.4)	(66.7)	(0)	(0)
BC[10,0]	1	0	0	22	1	0
	(4.2)	(0)	(0)	(91.7)	(4.2)	(0)
BC[5,5]	2	0	2	20	1	0
	(7.4)	(0)	(7.4)	(74.1)	(3.7)	(0)

Note: Reported are the actual actual number of observations. Percentages are in the parentheses.

In the experiment, one person from each pair will be randomly selected to be Player A and the other to be Player B. The players will interact in two stages: 1. The Negotiation Stage and 2. The Game. In the negotiation stage the players can form an agreement about how to play the game. Any agreement reached in the negotiation stage { *will or will not}* be enforced and the players {will have to play according to the agreement *or* be free to make any decisions} in the game that follows. The decisions in the negotiation stage will determine how much each of the players earns in the experiment.

We next describe first the game and then the negotiation stage that precedes it.

The Game

Player A moves first and chooses either IN or OUT by clicking a button labeled either "IN" or "OUT."

Player B moves second:

- If Player A chose OUT, then the game ends. *Player A receives* { *$5 or $10* } *and Player B receives* { *$5 or $0* } *.*
- If Player A chose IN, then *Player B splits $30 between the two of them*: Player B keeps $x and gives $30-x to Player A, choosing x such that $0 ■ x ■ $30.

The Negotiation Stage

Before the game is played the players can form an agreement about how to play the game. One player from each pair will be randomly selected to make the first proposal and the other player will be asked to respond to it. A proposal describes the choices of Player A and Player B in the game.

It could be:

Player A chooses OUT

or it could be

Player A chooses IN and Player B keeps \$x and gives \$30 – x to Player A.

The proposal is sent to the other player by clicking on the “Submit” button. The responding player observes the proposal and chooses one of the following three options:

- *Agree with the proposal* by clicking on the button “Agree.” In this case an agreement is formed and { *will or will not*} be enforced.
- *Make a counter-proposal* by clicking on the button “Make a counter-proposal.” This reverses the roles of the players in the negotiation. Now, the player who clicked this button makes a new proposal and sends it to the other player. The other player will then have the chance to respond by either agreeing with the proposal, or making a counter-proposal, or disagreeing.
- *Disagree and quit negotiating* by clicking on the button “Disagree and quit negotiating.” In this case no agreement is reached and negotiations terminate. Both players proceed to play the game.