Promises and Conventions -
A Theory of Pre-play Agreement*

Topi Miettinen†

February 2006

Abstract

Experiments suggest that communication increases the contribution
to public goods (Ledyard, 1995). There is also evidence that, when con­
templating a lie, people trade off their private benefit from the lie with
the harm it inflicts on others (Gneezy, 2005). We develop a theory of con­
ventions and bilateral pre-play negotiations that assumes the latter and
implies the former.

In symmetric games, agreements crucially depend on whether actions
are strategic complements or substitutes. With strategic substitutes, com­
mittance power tends to decrease in efficiency whereas the opposite may
be true with strategic complements. Also this finding is supported by
experimental evidence.

JEL Classification C72, C78, Z13

KEYWORDS: pre-play negotiation, communication, agreements, con­
ventions, norms, guilt

---

*I am grateful to the Yrjö Jahnsson Foundation for financial support. Also, I am very
grateful to Steffen Huck for advice, encouragement, suggestions, and discussions and also to
Philippe Jehiel on the same grounds. Finally, I would like to thank Martin Dufwenberg,
Daniel Friedman, Antonio Guarino, Klaus Helkama, Klaus Kultti, Mikko Leppämäki, Lisa
Myyry, Matthew Rabin, Francesco Squintani, Emma Tomainey, Joel Sobel, Pekka Stäskilahti,
Juuso Välimäki and the seminar participants at CEBR in Copenhagen, Game Theory Festival
at Stony Brook, Helsinki Center of Economic Research, University College London, Zeuthen
Workshop in Copenhagen. All errors are mine.

†Department of Economics, University College London, Gower Street WC1E 6BT and
RUESG, Dept of Economics, 00014 University of Helsinki. t.miettinen@ucl.ac.uk.
Contents

1 Introduction 3
1 Introduction

Ray and Cal have a magic pot and ten dollars each. Each dollar put into the pot gives $\frac{3}{4}$ dollars to both of them. Ray and Cal have to decide how many dollars to put into the pot and how many to keep to themselves. Ray figures that, whatever Cal puts into the pot, for each dollar he puts into the pot, he gets only $\frac{3}{4}$ dollars back and, hence, should put nothing into the pot.

Before they decide, they can talk to each other. They may agree on how many dollars each of them will put into the pot. The agreement is not binding. Yet, having talked to Cal for a while, he seems like nice guy to Ray. Ray starts to think that he would feel bad if he lied about how many dollars he will put into the pot. He also figures that Cal may well think similarly about him. Eventually, Ray and Cal agree on putting ten dollars each into the pot and neither violates the agreement.

Most people would think that the story above is vaguely plausible but doubt that such magic pots exist. An economist is certain about the existence of the magic pot, but has doubts whether people care about inflicting harm on the other by not doing as agreed.

Two findings in experimental economics give a reason to believe that the magic pots and the dislike to breach oral agreements are worth taking seriously: First, communication increases contributions in public good games\(^1\). Second, if people lie, they tend to dislike it; and the more harm they inflict on others by doing so, the more they seem to dislike it. This is shown by Gneezy (2005) and by several studies in social psychology. In public good games, agreeing to contribute more than one actually intends to contribute amounts to a lie which harms others. Thus, a theory that assumes the latter finding provides an explanation for the former finding.

This paper presents a theory of pre-play negotiations which assumes such a dislike to lie and shows that in an important class of symmetric games where actions are (weak) strategic complements (Bulow, Roberts and Klemperer, 1985), a symmetric efficient agreement can be made, if any. This class includes the public good provision game with constant returns to scale production technology and other team work and partnerships designs as well as Bertrand duopolies with imperfect substitutes. Yet, in symmetric games with strategic complements, such as the public good provision with a decreasing returns to scale technology,

\(^1\)See Ledyard (1995) for a review of experimental research on public goods. This result holds for public good games without a threshold. The evidence that communication would increase contributions in public good games with thresholds is more mixed - the increase in contributions is not always significant.
or the Cournot duopoly, there tends to be a conflict between the efficiency of the agreement and the incentives to respect it.

Public good experiments with communication lend strong support for our theory: Isaac and Walker (1988) adopt a constant returns to scale technology and find a strong positive effect of communication on efficiency. Average contribution levels are practically first-best efficient. Isaac, McCue and Plott (1985) adopt a decreasing returns to scale production technology. Despite the positive effect of communication on efficiency, they find that the average contribution levels are well below first-best efficient\(^2\).

Our theory considers bilateral pre-play negotiations in a wide array of strategic two-player interactions. The underlying game, the game that is played after the negotiations, can be any normal form game. We assign the guilt cost properties that experimental and narrative research in psychology and economics has discovered. We assume that the general principles that govern guilt are the same for all players. Players may differ only in their proneness to guilt, i.e. how much weight they put on the guilt cost. We abstract from the specificities of the negotiation protocol but assume that the outcome of such negotiation (agreement) is either an action profile of the underlying game or disagreement. Having agreed on a profile, a player who breaches may feel guilty, which lowers her utility.

Given a game and players’ proneness to guilt, each agreement maps the game into another game with the same strategy sets, but different payoffs. We are interested in which action profiles are agreeable, which action profiles can be enforced by guilt when players voluntarily enter into informal agreements to play accordingly. Also, we are interested in how agreeability is affected by changes in (1) the underlying game, (2) the agreement, and (3) players’ proneness to guilt.

Agreeability is defined in terms of incentive compatibility and individual rationality. An action profile is incentive compatible if neither player prefers breaching. That is, for any unilateral deviation from the profile, the guilt cost is larger than or equal to the underlying game benefit for the deviator. We call the difference between the underlying game benefit and the guilt cost the incentive to breach. An action profile is individually rational if it ensures that each player gets more than in her least preferred Nash equilibrium of the underlying game. Upon deciding whether to impose disagreement, each player acts as if she knew that doing so will imply coordination on her worst Nash equilibrium.

Which agreements are agreeable will depend crucially on the properties of the guilt cost. We adopt the following properties, which are based on stylized facts in research in social psychology and experimental economics\(^3\):

\{A\} Guilt costs are weakly increasing in the harm a player inflicts on his opponent by breaching an agreement.

\(^2\)These two studies are the only one’s that allow subjects to play repeatedly and learn about the game. Actually, Isaac and Walker (1988) have one design with constant returns to scale technology and another with decreasing returns to scale technology. With the former design, first best efficiency is reached whereas the latter falls short of first best.

\(^3\)In addition to their intuitive appeal, we present experimental evidence and psychological theory that supports these assumptions in section ??.
If the opponent breaches, then there is no guilt cost.

Guilt costs are weakly increasing in the player’s agreed payoff.

If no agreement is reached, there is no guilt cost.

Property {A} captures the idea that if my breaching the agreement causes my opponent to lose a toe, I do not suffer more than if my breaching the agreement causes my opponent to lose a leg. Gneezy (2005) finds strong support for property {A}: his experiments suggest that people trade off the benefits of lying against the harm that lying inflicts on the opponent. Property {B} is a no-sucker property: I will not feel guilty about breaching an agreement if my opponent breaches the agreement, too. According to property {C} an opponent’s generosity and kindness induce stronger guilt. Since there is guilt only if the opponent does not breach the agreement, the fact that the opponent respects and the fact that the agreed payoff is high indicate that the opponent is kind and generous. Hence, breaching the agreement and not reciprocating this will induce stronger guilt than if the agreement had been less generous. Properties {B} and {C} render guilt reciprocal. Property {D} expresses the idea that if nothing is promised then there can be no guilt of breaching.

Crucial for our finding in games with strategic complements and substitutes and an interesting result in its own right is that, in games where actions are ordered and the payoff is concave in each of the two actions, checking that a marginal deviation from the agreement does not pay off is necessary and sufficient for incentive compatibility.

Further towards our main conclusion, we find unambiguous effects on the incentive to breach when the terms of the agreement are altered (if the agreement is agreeable in the first place): in symmetric games with strategic complements, changing either agreed action so as to improve a player’s agreed payoff decreases her marginal incentive to breach. These effects are quite natural and intuitive: if the terms of the agreement are better for me, I have a lower incentive to breach. Yet, the result does not hold generally.

In symmetric games with strategic substitutes, as far as changes in player’s own action are concerned, the player’s payoff and her incentives to respect agreements are still naturally aligned. Yet, changing the opponent’s agreed action implies quite the opposite effect: the marginal benefit increases and the marginal harm on the opponent decreases when the opponent’s action is changed so as to improve player’s payoff. This is the source of our result, identifying a conflict between efficiency and incentives in symmetric games with strategic substitutes, such as the standard Cournot duopoly or public good provision with a decreasing returns to scale production technology.

We also describe the agreeable set in a more general class of games and characterize the smallest and largest such set: Nash equilibria are always agreeable and nothing but Nash equilibria are agreeable for players with no proneness to guilt. Yet, a player who is sufficiently prone to guilt can agree on any individually rational profile that she cannot alone Pareto-improve and strictly benefit herself.
The paper is organized as follows. Section ?? presents related literature in economics and psychology. Section ?? presents the model. Section ?? studies a public good game. Section ?? presents general results and section ?? studies games with ordered strategy spaces. Section ?? considers a Cournot duopoly example. Section ?? concludes and discusses some further research problems.