http://www.aeaweb.org/articles.php?doi=10.1257/aer.98.4.1629

Commitment and Conflict in Bilateral Bargaining

By Tore Ellingsen and Topi Miettinen*

If negotiators can write binding contracts, why is there ever costly disagreement in bilateral bargaining? Existing formal theory emphasizes two main possibilities. Disagreement arises because of incomplete information or because of irrational negotiators. Thomas C. Schelling (1956, 1960, 1966) proposes a third reason.¹ Rational negotiators may attempt to increase their share of the available surplus by visibly committing themselves to an aggressive bargaining stance and thereby forcing concessions from an uncommitted opponent. If both negotiators simultaneously make such strategic commitments, there is conflict.

Vincent P. Crawford (1982) pursues Schelling’s argument in a formal model in which negotiators decide simultaneously whether to attempt commitment, and if so, which stance to attempt commitment to.² In the case of symmetric information, Crawford concludes that there are often many equilibria, some of which are both efficient and equitable.³ When both negotiators can credibly commit not to accept any less than half the surplus, there is also typically an equilibrium in which they do so. Such compromise equilibria fail to exist, however, if commitments are weak—in the sense that they are likely to be revokable at low cost. For example, if any commitment sticks with probability q and is freely revokable with probability 1 − q, then the unique equilibrium entails commitment attempts by both negotiators if q is sufficiently small.⁴ The intuition is that if the opponent’s commitment is revokable with high probability, the best strategy is to try to exploit the opponent’s expected weakness by making an aggressive commitment. When both negotiators succeed in making their commitments stick, there is conflict.

An objection to this theory of conflict is that negotiators often have access to quite efficient commitment technologies. For example, a political leader who publicly and emotionally states that she will never yield a piece of territory to another country may find it very difficult to

¹ Schelling’s (1956) seminal essay on bargaining is reprinted as Chapter 2 of Strategy of Conflict (Schelling 1960). The book’s other chapters, and especially Appendix B, contain substantial additional thoughts on commitments in bilateral bargaining. In Arms and Influence (Schelling 1966), the ideas are used to shed light on international relations.
² Crawford (1982, 610–11) argues, like John F. Nash (1953), Schelling (1960, 70; Appendix B), and John C. Harsanyi and Reinhard Selten (1988, 23–26), that such simultaneous-move bargaining games can capture the dynamic give-and-take of bargaining and the role of expectations in determining outcomes. See also Crawford (1991).
³ Subsequent work by Abhinay Muthoo (1996) has investigated ways to obtain unique equilibrium selection in a related model when the cost of revoking a commitment is not prohibitive. Muthoo obtains an efficient compromise solution as the unique outcome.
⁴ This is a special case of Crawford’s model, to which he devotes considerable attention. Note, however, that Crawford also studies cases in which it is costly to revoke commitments and where players have private information about the costs of revoking. Subsequent literature has tended to focus on issues that arise in the latter asymmetric information case. (See, especially, the analysis of dynamic bargaining games with unobservable commitments by Shinsuke Kambe (1999), Dilip Abreu and Faruk Gul (2000), and Olivier Compte and Philippe Jehiel (2002).) Here, we focus exclusively on the case of observable commitments.
revoke this commitment unless something unexpected happens that can serve as an acceptable excuse for doing so. Since both negotiators prefer strong commitments to weak ones, real negotiations may take place with values of \( q \) that are high enough for efficient equilibria to exist.

In this paper we show that incompatible commitments pose a severe problem even when commitment technologies are highly credible, that is, when \( q \) is close to one. We demonstrate the point in two ways.

Our first approach is to modify Crawford’s model slightly by introducing a cost of commitment, \( c > 0 \). In the modified model with \( q < 1 \), we show that only the incompatible commitment equilibrium exists as \( c \) goes to zero. Thus, if the commitment cost is sufficiently small relative to the size of the contested surplus, there is conflict with probability \( q^2 \) even for large \( q \). In fact, disagreement is not only the unique Nash equilibrium outcome, but also the unique outcome to survive iterated elimination of strictly dominated strategies. When \( q = 1 \), efficient compromises remain unattainable, but in this case there are also equilibria in which one player gets all the surplus (in addition to the conflicting commitment outcome).

Our second approach is to retain all Crawford’s assumptions, including \( c = 0 \), while utilizing a stronger solution concept than Nash equilibrium. To be precise, we show that the outcomes above are the unique outcomes to satisfy the requirement that negotiators use only strategies that are iteratively weakly undominated.

To understand why efficient compromise outcomes fail to exist, consider the case in which both negotiators commit to take exactly half the surplus. If it is costly to make such a commitment, a negotiator is better off being flexible, as this yields the same share of the surplus. At the same time, it cannot be an equilibrium that both negotiators remain flexible, since one negotiator is then better off by making an aggressive commitment. If commitment costs are zero, essentially the same logic applies, except now moderate commitments are only iteratively weakly dominated.

These findings turn conventional wisdom on its head. In most models of bargaining under symmetric information, equilibria have a strong tendency to yield efficient outcomes. In our analysis, by contrast, conflict is frequently the unique equilibrium outcome.

I. Certain Commitment

There are two negotiators, henceforth called players. Players are indexed \( i = 1, 2 \) and bargain over a surplus of size 1. The size of the surplus and the rationality of the players are common knowledge.

In the first stage, each player \( i \) chooses, simultaneously with the other, either to commit to some demand \( s_i \in [0, 1] \) or to wait and remain uncommitted. Let \( w \) denote the waiting strategy. Without loss of generality, the cost of remaining flexible is normalized to zero. Committing entails a commitment cost \( c \geq 0 \). In the second stage, two uncommitted players engage in bargaining. Let \( \beta_i > 0 \) be player \( i \)'s share if both players are uncommitted in the second stage. We assume that two flexible players are able to coordinate on an efficient outcome, so that \( \beta_1 + \beta_2 = 1 \). Since many explicit models of noncooperative bargaining under perfect information result in such an interior solution, this can be taken as a reduced form of an (unmodelled) ensuing bargaining game. Without loss of generality, let \( \beta_1 \geq \beta_2 \).

In the second stage, a committed player cannot revoke her demand. An uncommitted opponent thus observes the opponent’s first-stage choice and best responds by demanding the residual share \( 1 - s_i \). If both players commit, the outcome depends on whether the commitments are compatible or not. If the commitments are compatible, \( s_1 + s_2 \leq 1 \), we assume that each player \( i \) gets a utility of at least \( s_i - c \) and at most \( 1 - s_i - c \). That is, we make no restriction on the
allocation of any unclaimed surplus $1 - s_i - s_j$. If the commitments are incompatible, $s_1 + s_2 > 1$, there is disagreement and both players get utility $-c$.

To summarize, each player has the set of pure strategies $S = [0, 1] \cup \{w\}$. Letting $x_i(s_i, s_j) \in [s_i, 1 - s_j]$ denote player $i$'s share of the surplus in the compatible commitments case, the payoff of player $i$ depends on these strategies as follows:

$$u_i(s_i, s_j) = \begin{cases} 
  x_i(s_i, s_j) - c & \text{if } s_i + s_j \leq 1; \\
  s_i - c & \text{if } s_j = w \text{ and } s_i \neq w; \\
  -c & \text{if } s_i + s_j > 1; \\
  1 - s_j & \text{if } s_i = w \text{ and } s_j \neq w; \\
  \beta_i & \text{if } s_i = w = s_j.
\end{cases}$$

The set of mixed strategies is the set of probability distributions on $S$. We write a mixed strategy of player $i$ as $\sigma_i$. Let $p_i(s)$ denote the associated probability that player $i$ plays the pure strategy $s$.

The following lemma is the key to our main result. It says that if commitment is more costly than flexibility, then nothing but a greedy demand of $1$, and waiting, $w$, is iteratively strictly undominated and thus played with a positive probability in any Nash equilibrium.

**Lemma 1:** Suppose $c > 0$. Then, for each player, only $1$ and $w$ are iteratively strictly undominated strategies.

**Proof:**

Observe first that player 1 strictly prefers $w$ to any $s_1 \in [0, \beta_1]$, and player 2 strictly prefers $w$ to any $s_2 \in [0, 1 - \beta_1]$; the commitment strategy gives player 1 a payoff $s_i - c < \beta_i$ when player $j$ chooses $w$. It gives at most $1 - s_i - c < 1 - s_j$ when player $j$ chooses a compatible commitment $s_j \in [0, 1 - s_j]$. Finally, it gives $-c < 1 - s_j$ when player $j$ chooses an incompatible commitment.

After these strategies are eliminated, it is easy to check that player 1 strategies $s_j \in (\beta_1, 1)$ are strictly dominated by the mixed strategy $\sigma_1 = (p_1(1) = s_1, p_1(w) = 1 - s_1)$. If player 2 plays $w$, player 1’s payoff to the pure strategy $s_1$ is $s_1 - c$; whereas, under the mixed strategy $\sigma_1$, he gets $s_1 + (1 - s_1) \beta_1 - s_1 c$. If player 2 plays $s_2 \in (1 - \beta_1, 1)$, then the payoff to a pure commitment strategy in $s_j \in (\beta_1, 1)$ is $-c$, whereas the payoff to the mixed strategy is $(1 - s_1)(1 - s_2) - s_1 c$ where the latter is clearly strictly greater.

Likewise, player 2 strategies $s_2 \in (1 - \beta_1, 1)$ are dominated by the mixed strategy $(p_2(1) = s_2, p_2(w) = 1 - s_2)$. Thus, only $w$ and $1$ are iteratively undominated.

The proof demonstrates that flexibility dominates modest commitments, and that once modest commitments are eliminated, there is no rationale for any other interior commitment either; extreme commitment or flexibility (or a probabilistic combination thereof) will be better.

The iterative elimination of strictly dominated strategies ultimately induces a game of Chicken. In this game, neither $(1, 1)$ nor $(w, w)$ constitutes an equilibrium. If $c < \beta_1$, $s_i = 1$ is the unique

---

5 Our specification of the compatible commitments outcome nests many special cases. If players always obtain exactly $s_i - c$ under compatible commitments, we have the Nash demand game allocation. This specification is unintuitive, since unclaimed surplus is left on the negotiation table. If the full surplus is allocated according to relative claims, we have, instead, the special case considered by Ellingsen (1997).
best response to \( s_j = w \), and \( s_i = w \) is the unique best response to \( s_j = 1 \). There is thus a unique pair of asymmetric pure strategy equilibria. In these equilibria, one player commits to demanding all the surplus and the other player waits. There is no symmetric pure strategy equilibrium, however. In addition to the asymmetric pure strategy equilibria, there is one in mixed strategies. In a mixed strategy equilibrium, the expected payoff from waiting, which is \( 1 - p_j(1) - c \), must equal the expected payoff from waiting, which is \( (1 - p_j(1))\beta_i \). Clearly, for \( \beta_1 = \beta_2 \), this mixed strategy equilibrium is symmetric.

**Proposition 1:** Suppose \( c \in (0, \beta_2) \): (i) for any pair \( (\beta_1, \beta_2) \), there is a mixed strategy equilibrium, \( p_1(1) = (\beta_1 - c)/\beta_i \) and \( p_1(w) = c/\beta_i \); (ii) in the unique pair of pure strategy equilibria, either \( p_1(1) = 1 \) and \( p_2(w) = 1 \) or \( p_2(1) = 1 \) and \( p_1(w) = 1 \).

As noted already by Crawford (1982), there is a plethora of efficient equilibria when \( c = 0 \). The discontinuity at \( c = 0 \) may seem strange. If we refine the set of Nash equilibria, however, the discontinuity vanishes.

**Lemma 2:** Suppose \( c = 0 \). Then, for each player, only 1 and \( w \) are iteratively weakly undominated strategies.

**Proof:**

Step (i): Any first stage commitment \( s_i \leq \beta_i \) is weakly dominated by \( w \), because \( w \) earns at least as much against all opponent’s commitments \( s_j \neq 1 - s_i \) and the same against \( s_j = 1 - s_i \), and either more or the same against \( s_j = w \).

Step (ii): Given that opponent strategies \( s_i \leq \beta_i \) are likewise eliminated, any commitment strategy \( s_i \in (\beta_i, 1) \) is now weakly dominated by 1, since 1 earns the same as \( s_i \) for all weakly undominated opponent strategies except \( w \); in the latter case 1 earns more.

With the lemma in hand, it is trivial to prove our next result.**

**Proposition 2:** Suppose \( c = 0 \). For any pair \( (\beta_1, \beta_2) \), only three subgame perfect Nash equilibrium outcomes are consistent with two rounds of elimination of weakly dominated strategies. These equilibria are (i) \( p_1(1) = p_2(1) = 1 \); (ii) \( p_1(1) = 1 \) and \( p_2(w) = 1 \); and (iii) \( p_2(1) = 1 \) and \( p_1(w) = 1 \).

Our mixed strategy equilibrium can be seen as a cousin of Proposition 2 in Ellingsen (1997). Ellingsen also studies the trade-off between commitment and flexibility in the Nash Demand game, albeit in a single-population evolutionary model with observable strategies and zero commitment costs. His Proposition 2 considers a case in which the size of the pie is uncertain, and commitments are nominal. In that case, a bilateral commitment to “fair” demand, such as half the normal surplus, will entail conflict whenever the surplus is smaller than normal. The flexible strategy then does better than the committed fair strategy, and ends up coexisting with the aggressive committed strategy. Ellingsen’s result hinges crucially on the assumption that the size of the surplus is uncertain when the commitment is made, and that it is impossible to commit to a relative share of the surplus. We make neither of these assumptions.

In another closely related paper on the Nash Demand game, Werner Güth, Klaus Ritzberger, and Eric van Damme (2004) assume that there is some small uncertainty as to the actual size

---

6 For completeness, let us mention that, if commitment would be cheaper than remaining flexible, many other equilibria are sustainable and there are no weakly dominated strategies.
of the surplus, and let negotiators choose their demands either before or after the uncertainty resolves. They show that there are only two strict Nash equilibria. In each of these equilibria, one negotiator demands almost all of the surplus before the uncertainty is resolved. The other waits until the size of the pie is known and demands the remainder. In their model, as in Ellingsen’s, commitment is costly because of the inability to adapt the demand perfectly to the size of the surplus. However, by only reporting asymmetric strict equilibria, they predict an efficient outcome. It is worth pointing out that the asymmetric equilibria are strict only because of the uncertain size of the surplus. With full certainty of the surplus size, Güth, Ritzberger, and van Damme are back to the vast multiplicity of equilibria.

II. Uncertain Commitment

In this section, we maintain the assumption that players can attempt commitment at a cost $c$, but relax the assumption that the commitment succeeds with probability 1. Define $q_1 = q_2 = q$ as the probability (independent of the size of commitment) of success. We consider only the case $0 < c \leq q(1 - \beta_i)$ in which commitment is potentially attractive.

As a backdrop, consider first the argument of Crawford (1982, 620–2) that commitments are tempting. Suppose $s_i = \frac{1}{2}$ as the probability (independent of the size of commitment) of success. We consider only the case $0 < c \leq q(1 - \beta_i)$ in which commitment is potentially attractive. As we shall now see, the efficient equilibria that exist when $q \in (\frac{1}{2}, 1)$ are highly sensitive to the assumption $c = 0$. More precisely, when $c$ is positive but sufficiently small, the efficient equilibria vanish.

PROPOSITION 3: (i) If $0 < c < q(1 - q)(1 - \beta_i)$, then for each player $i$ only $s_i = 1$ is an iteratively strictly undominated strategy. (ii) If $0 < q(1 - q)(1 - \beta_i) \leq c \leq q(1 - \beta_i)$, then only $1$ and $2$ are iteratively strictly undominated strategies.

PROOF:

Observe first that player $i$ strictly prefers $w$ to any $s_i \in [0, \beta_i]$: the commitment strategy $s_i$ yields $q_s + (1 - q) s_i + c < \beta_i$, when player $j$ chooses $s_j$. It yields $q_s + (1 - q) s_j + (1 - q) \beta_i c < q(1 - s_j) + (1 - q) \beta_i$ when player $j$ chooses a compatible commitment in $[0, 1 - s_j]$, and finally it yields $q(1 - q) s_i + (1 - q) q(1 - s_j) + (1 - q)^2 \beta_i c < q(1 - s_j) + (1 - q) \beta_i$ when player $j$ chooses an incompatible demand.

After these commitment strategies are eliminated, it is easy to check that strategies $s_1 \in (\beta_1, 1)$ are strictly dominated by the mixed strategy $\sigma_1 = (p_1(1) = s_1, p_1(w) = 1 - s_1)$: if player 2 plays $w$, player 1’s payoff to a pure strategy in $s_1 \in (\beta_1, 1)$ is $q s_1 + (1 - q) s_i + \beta_i c$; whereas under the mixed strategy $\sigma_1$ he gets $q s_1 + (1 - q s_1) \beta_i < s_1 \beta_i$. If player 2 plays $s_2 \in (1 - \beta_2, 1)$, the payoff to a pure commitment strategy in $s_1 \in (\beta_1, 1)$ is $q(1 - q) s_1 + (1 - q) q(1 - s_2) + (1 - q)^2 \beta_i c$, whereas the payoff to the mixed strategy is $q s_1(1 - q) + (1 - q s_1) q(1 - s_2) + (1 - q s_1)(1 - q) \beta_i s_2 \beta_i c$, which is strictly greater. Likewise, player 2 strategies $s_2 \in (1 - \beta_2, 1)$ are dominated by the mixed strategy $(p_2(1) = s_2, p_2(w) = 1 - s_2)$.

Suppose now that $c < q(1 - q)(1 - \beta_i)$. Since all other strategies are eliminated, it is easy to check that $s_i = 1$ dominates $w$: if the opponent waits, $s_i = 1$ gives $q + (1 - q) \beta_i c$ and $w$ gives $\beta_i$. Commitment is the better strategy, since $q(1 - \beta_i) > q(1 - q)(1 - \beta_i) > c$. If the opponent
plays \( s_j = 1 \), playing \( s_i = 1 \) yields \( q(1 - q) + (1 - q)(1 - q)\beta_i - c \) while \( w \) yields \( (1 - q)\beta_i \). Since \( c < q(1 - q)(1 - \beta_i) \), commitment is again better. Thus, part (i) is established. On the other hand, if \( c \geq q(1 - q)(1 - \beta_i) \), \( s_i = w \) is a best response to \( s_j = 1 \) but \( s_i = 1 \) is a best response to \( s_j = w \). This establishes part (ii).

The most remarkable feature of the result is that, when \( c \) is sufficiently small, conflict occurs with probability \( q^2 \) in a unique equilibrium even as \( q \) is arbitrarily close to 1. While the probability of conflict in the best equilibrium based on Crawford’s analysis is at most \( (1/2)^2 = 1/4 \), in the slightly modified game the best equilibrium may instead have conflict with almost full certainty.

Since we normalize the size of the total surplus to one, it is the ratio of the commitment cost to the size of the surplus that matters for the likelihood of conflict. If the cost of making a commitment is independent of the surplus to be shared, conflict is more likely when the surplus is large. More generally, if commitment costs increase less than proportionally with the size of the surplus, our model is consistent with the notion that some issues are too small to be fighting about. Conversely, any sufficiently big issue invites conflict.

From a technical point of view, making commitment somewhat uncertain does not affect the first two iterations in Lemma 1. Yet, it changes the strategic aspects of the remaining \( 2 \times 2 \) game. With certain commitment, this was a game of Chicken; with uncertain commitment and low cost of attempting, it becomes a Prisoner’s Dilemma: attempting commitment is a strictly dominant strategy.\(^7\)

Finally, let us again consider the role of equilibrium refinement in the limit case \( c = 0 \).

**Proposition 4:** Suppose \( c = 0 \). Then, for each player \( i \), only \( s_i = 1 \) is an iteratively weakly undominated strategy.

**Proof:**

**Step (i).** Any first-stage commitment \( s_i \leq \beta_i \) is weakly dominated by \( w \). The commitment strategy yields \( qs_i + (1 - q)\beta_i \leq \beta_i \) when player \( j \) chooses \( w \). It yields at most \( q^2(1 - s_i) + q(1 - q) s_i + (1 - q) q(1 - s_j) + (1 - q)^2 \beta_i \leq q(1 - s_i) + (1 - q) \beta_i \) when player \( j \) chooses a compatible commitment, and finally it yields \( q(1 - q) s_i + (1 - q) q(1 - s_j) + (1 - q)^2 \beta_i < q(1 - s_j) + (1 - q) \beta_i \) when player \( j \) chooses an incompatible commitment.

**Step (ii).** After opponent strategies \( s_j \leq \beta_j \) are likewise eliminated, any commitment strategy \( s_i \in (\beta_i, 1) \) is now weakly dominated by 1, since 1 earns the same as \( s_i \) for all weakly undominated opponent strategies except \( w \); in the latter case, 1 earns more. Finally, when all but \( s_i = 1 \) and \( s_j = w \) are eliminated, the former dominates the latter: if \( s_j = 1 \), then \( s_i = 1 \) yields \( q(1 - q) + (1 - q)^2 \beta_i \), whereas \( s_i = w \) yields \( (1 - q) \beta_i \), which is less.

**III. Final Remarks**

We have reexamined the problem of observable commitments in bargaining, first studied by Schelling (1956) and later formalized by Crawford (1982). Our benchmark is the special case of Crawford’s model in which commitment sticks with probability \( q \) and is freely revokable with probability \( 1 - q \). We show that either a slight modification of the model (the introduction of a

\(^7\) This result bears a relation to H. Peyton Young’s (1993) insight that a small chance of meeting a bargainer from one’s own population destabilizes asymmetric stable states. Here, the small uncertainty of failure of commitment plays a similar role as the probability of meeting a player from one’s own population.
positive cost of commitment) or a strengthening of the solution concept suffice to greatly alter the results. The new results potentially shed light on two types of phenomena.

The first phenomenon is the infrequency of side-payments. Although we lack systematic evidence, it seems to us that side-payments are less frequent than suggested by the compromise outcomes favored by most other bargaining models. Consider, for example, the case in which an unpleasant job can be carried out either by person A or by person B, but not by both of them (say, the job of being department chairperson). The compromise solution is that whoever takes on the job ought to be compensated by the other through a side-payment. Our model, like Schelling’s intuitive analysis, implies that such side-payments should not occur unless commitment costs are prohibitively high.

The second phenomenon is conflicts. If it is cheap to make commitments but impossible to make such commitments stick with probability 1, the model predicts that there is stalemate due to conflicting commitments. Thus, here we have a simple explanation of conflict that invokes neither asymmetric information nor exogenous constraints on side-payments. If anything, the model produces incompatible commitments all too readily.

Our analysis leaves out several considerations that may be important in practice. For example, the model allows commitments to be made in a single period only. Two natural avenues along which the analysis may be profitably extended are: allowing more time periods, and allowing (random) delays in the communication of commitments. After more than 50 years, formal game theory is still catching up with Schelling’s analysis of commitment tactics in bargaining.

REFERENCES