# Web Appendix:

## Coalition Formation with Switching Options

***M*** is the formateur and can bargain with ***L*** or ***H*** with payoffs set out in this note. In the absence of the possibility of ***M*** being able to threaten credibly to switch partners in a Rubinstein alternating offer bargaining game, if the payoff to ***M*** of choosing ***L*** is greater than that of choosing ***H***, ***M*** will choose to bargain with ***L*** (and will broadly split the difference with ***L***). This sharply contrasts with ***M*** being able to bargain with ***L*** or ***H*** on a take-it-or-leave-it basis, which produces a broadly median voter outcome. A criticism which has been made of the simple Rubinstein approach is that it is not robust to the possibility of switching. But we show in this note that no matter many times ***M*** is allowed to switch partners after ***M*** rejected an offer by the current partner the Rubinstein result holds.

Let $U_I(\sigma_I^J)$ be the utility party ***I*** gets from ***J***'s acceptance of ***I***'s offer of $\sigma_I^J$. We first define the Rubinstein unique SGPE in the alternate offer bargaining games in which there is no outside (or switching) option. In the ***LM*** game we have:

(1)
$$\begin{aligned} U_L(\bar{\sigma}_M^L) &= \delta U_L(\bar{\sigma}_L^M) \\ U_M(\bar{\sigma}_L^M) &= \delta U_M(\bar{\sigma}_M^L) \end{aligned}$$

Note that the initial offer that ***M*** makes to ***L*** is $\bar{\sigma}_M^L$, and this is accepted by ***L***. Thus the value to ***M*** of opening bargaining with ***L*** is $U_M(\bar{\sigma}_M^L)$.

In the ***HM*** game we have:

(2)
$$\begin{aligned} U_H(\bar{\sigma}_M^H) &= \delta U_H(\bar{\sigma}_H^M) \\ U_M(\bar{\sigma}_H^M) &= \delta U_M(\bar{\sigma}_M^H) \end{aligned}$$

Hence the initial offer that ***M*** makes to ***H*** is $\bar{\sigma}_M^H$, and this is accepted by ***H***. Thus the value to ***M*** of opening bargaining with ***H*** is $U_M(\bar{\sigma}_M^H)$.

Because the model implies that ***M*** gets more if ***H*** is excluded from the coalition than if ***L*** is excluded, when there is no possibility of switching we have that:

(3) $$U_M(\bar{\sigma}_M^L) > U_M(\bar{\sigma}_M^H)$$

i.e. in bargaining without switching options ***M*** has a higher payoff from bargaining with ***L*** than from bargaining with ***H***.

The "outside option principle" (see, e.g. Osborne and Rubinstein [1994], 7.4.3, p 128) states that if the value to ***M*** of switching to bargaining with say ***H*** after rejecting an offer of ***L*** is $\bar{U}_M^{MH}$, then

(i) if $\bar{U}_M^{MH} > U_M(\bar{\sigma}_M^L)$ ***L***'s SGPE offer to ***M*** is $\delta\bar{U}_M^{MH}$ and (ii) if $\bar{U}_M^{MH} \le U_M(\bar{\sigma}_M^L)$ this has no effect on the ***LM*** game and ***L*** continues to offer $U_M(\bar{\sigma}_L^M) = \delta U_M(\bar{\sigma}_M^L)$. We call case (i) the case of a "binding" outside option and (ii) a "non-binding" option.

If there is a binding option in the ***LM*** game, ***M***'s SGP offer to ***L***, $\tilde{\sigma}_M^L$, is given by the solution of

(4) $$\begin{aligned} U_L(\tilde{\sigma}_M^L) &= \delta U_L(\tilde{\sigma}_L^M) \\ U_M(\tilde{\sigma}_L^M) &= \delta\bar{U}_M^{MH} \end{aligned}$$

If there is a binding option in the ***MH*** game, ***M***'s SGP offer to ***H***, $\tilde{\sigma}_M^H$, is given by the solution of

(5) $$\begin{aligned} U_H(\tilde{\sigma}_M^H) &= \delta U_H(\tilde{\sigma}_H^M) \\ U_M(\tilde{\sigma}_H^M) &= \delta\bar{U}_M^{ML} \end{aligned}$$

**Lemma 1.** In the Switching Game, if $U_M(\bar{\sigma}_M^L) > U_M(\bar{\sigma}_M^H)$ then $\bar{U}_M^{ML}$ is binding on the ***MH*** game and $\bar{U}_M^{MH}$ is not binding on the ***LM*** game.

> *Proof:* It cannot that $\bar{U}_M^{ML}$ is binding on the ***MH*** game and $\bar{U}_M^{MH}$ is binding on the ***LM*** game, or equivalently that both are non-binding. In those cases $\bar{U}_M^{ML} = U_M(\bar{\sigma}_M^L) = \bar{U}_M^{MH} = U_M(\bar{\sigma}_M^H)$ which is contradicted by $U_M(\bar{\sigma}_M^L) > U_M(\bar{\sigma}_M^H)$. So we need to show it is not the case that $\bar{U}_M^{MH} = U_M(\bar{\sigma}_M^H)$ is binding on the ***LM*** game. If it is $\bar{U}_M^{MH} = U_M(\bar{\sigma}_M^H) > U_M(\bar{\sigma}_M^L)$ which is also contradicted by $U_M(\bar{\sigma}_M^L) > U_M(\bar{\sigma}_M^H)$.

We now show that if ***M*** can choose whom to bargain with, given the possibility of switching, ***M*** will choose to bargain with ***L*** and not switch. The payoff to the ***LM*** bargain is $U_M(\bar{\sigma}_M^L)$. The payoff to the ***MH*** bargain is $U_M(\tilde{\sigma}_M^H)$ and we need to establish that $U_M(\bar{\sigma}_M^L) > U_M(\tilde{\sigma}_M^H)$.

**Lemma 2.** $U_M(\bar{\sigma}_M^L) > U_M(\tilde{\sigma}_M^H)$.

*Proof:* Subtract the second equation in (2) from the second equation in (5), and using the positive monoticity of $U_M(\sigma)$

$$\left[U_M(\tilde{\sigma}_H^M) - U_M(\bar{\sigma}_M^H)\right] = \delta\left[\bar{U}_M^{ML} - U_M(\bar{\sigma}_H^M)\right] > 0$$
$$\rightarrow U_M(\tilde{\sigma}_H^M) > U_M(\bar{\sigma}_M^H) \ \rightarrow \tilde{\sigma}_H^M > \bar{\sigma}_M^H$$

Since $U_H(\sigma)$ is monotonically decreasing, $\tilde{\sigma}_H^M > \bar{\sigma}_M^H \rightarrow U_H(\tilde{\sigma}_H^M) < U_H(\bar{\sigma}_H^M)$, so if we subtract the first equation in (2) from the first equation in (5)

$$\left[U_H(\tilde{\sigma}_M^H) - U_H(\bar{\sigma}_M^H)\right] = \delta\left[U_H(\tilde{\sigma}_H^M) - U_H(\bar{\sigma}_H^M)\right] < 0$$

which implies $\tilde{\sigma}_M^H > \bar{\sigma}_M^H \ \rightarrow U_M(\bar{\sigma}_M^H) < U_M(\tilde{\sigma}_M^H) = \bar{U}_M^{MH} < \bar{U}_M^{LM} = U_M(\bar{\sigma}_M^L)$. Therefore $U_M(\tilde{\sigma}_M^H) = \bar{U}_M^{MH} < \bar{U}_M^{LH} = U_M(\bar{\sigma}_M^L)$.

A corollary of this is that $U_M(\bar{\sigma}_M^H) < U_M(\tilde{\sigma}_M^H) = \bar{U}_M^{MH}$. In other words, the possibility that ***M*** can switch to ***LM*** from ***MH*** means that ***H*** can make a higher credible offer to ***M*** than in the absence of the possibility of switching, but not high enough to tempt ***M*** away from ***L*** (since it is still the case that $U_M(\tilde{\sigma}_M^H) = \bar{U}_M^{MH} < U_M(\bar{\sigma}_M^L)$).

The main result is stated in the following proposition:

**Proposition.** Assume that (a) ***M*** can choose with whom to bargain at the start of the coalition choosing process, and (b) on any finite number of occasions $N = 1,..,\bar{N} < \infty$ after ***M*** has rejected a counter-offer from its current bargaining partner, that (c) ***M*** makes the first offer in any new bargaining game, and that it is otherwise a standard Rubinstein infinitely repeated alternative offer extensive game. Then if $U_M(\bar{\sigma}_M^H) < U_M(\bar{\sigma}_M^L)$, ***M*** will always choose to bargain with ***L***, will offer ***L*** $\bar{\sigma}_M^L$ on the first move and ***L***will accept it.

*Proof:* Assume ***M*** has switched partners the penultimate maximum number, $\bar{N}$ - 1, of times. If it now rejects ***I***'s offer, it will choose to bargain with ***L*** as its $\bar{N}$ th partner since no further switching is possible and since $U_M(\bar{\sigma}_M^H) < U_M(\bar{\sigma}_M^L)$. If it goes back one partner, it will also choose ***L*** since $U_M(\tilde{\sigma}_M^H) < U_M(\bar{\sigma}_M^L)$; ***H*** can make a better offer now but not good enough. If it goes back one further, ***M*** faces the same choice at $\bar{N}-2$ as at $\bar{N}-1$. This is because the payoff from choosing ***L*** is always the same, $U_M(\bar{\sigma}_M^L)$, since there is no outside option in this case; and similarly the payoff to choosing ***H*** is always the same since the payoff , $U_M(\tilde{\sigma}_M^H)$, is derived as before from

$$\begin{aligned}U_H(\tilde{\sigma}_M^H) &= \delta U_H(\tilde{\sigma}_H^M)\\ U_M(\tilde{\sigma}_H^M) &= \delta \bar{U}_M^{ML} = \delta U_M(\bar{\sigma}_M^L)\end{aligned}$$

Hence for all *N*, $1 \leq N < \bar{N}$, ***M*** will choose ***L*** and ***L*** will accept ***M***'s offer of $\bar{\sigma}_M^L$ at once. Since this is true for all positive feasible numbers of switches $\bar{N}$, the ability to switch coalition partners does not alter the result that will choose ***L*** and ***L*** will accept ***M***'s offer of $\bar{\sigma}_M^L$ at once at the start of the coalition formation process.

The intuition for the result is that offers that deviate from the Rubinstein solution are not credible unless the outside option is binding. For example, if ***H*** offers ***M*** a better deal than ***M*** is currently getting from an *LM* coalition, such an offer would not be time-consistent because once ***M*** leaves ***L*** for a new coalition with ***H*** there is a cost of moving back to ***L***. This cost allows *H* to renege on its original offer and give ***M*** what it was already getting from the *LM* coalition, *discounted* by the discount factor squared. This logic applies no matter how many times ***M*** has switched in the past, so ***M*** will never switch.