# Appendix to article “Contested Multilateralism” as Credible Signaling: How Strategic Inconsistency Can Induce Cooperation among States

In the body of the paper, we develop two theoretical propositions based on the spatial model embodied in Figures 1 and 2 and on the verbal reasoning presented in Sections 3 and 4. We also note that the necessarily static spatial model does not capture the temporal dynamics that form part of our theorizing. The core purpose of the computer simulation presented in this appendix is to probe the validity of our theoretical reasoning in particular with regard to the effects of time on the cooperation gains reaped by states. This analysis thus elaborates on how variation in the key timing parameters we hint at towards the end of Sections 3 and 4, respectively, influences the effects of CM on the gains that states can reap from cooperation.

The appendix is structured as follows. *First*, we present the setup of the computer simulation. *Second*, we show core evidence pertinent to our Propositions 1 and 2 in the dynamic setting we model here. *Third*, we explicitly model the effects of time on the gains or losses due to CM. *Fourth*, we evaluate the robustness of our quantitative results. *Fifth*, we probe the relevance of actors’ patience for our results.

## 1. Simulation Setup

The core of our approach to assessing the effects of CM on patterns of cooperation in a dynamic perspective involves comparing actors’ gains under two competing scenarios; one based on the challenger’s choice for CM, and one without it (the baseline scenario). In the spatial model, we can specify three core institutional outcomes of interest: 1) the location of the initial status quo; 2) the bi-institutional situation, with competing sets of rules, after CM has been exercised; and 3) the new state of inter-institutional complementarity after post-CM inter-institutional adjustment has successfully occurred. Based on the utility functions of both actors, we are able to calculate the payoffs actors would receive from each of these institutional outcomes. Because the path of CM outlined in Sections 3 and 4, as well as the baseline path without CM, represent a move of the system between these institutional outcomes, we can compute each actor’s gains under the two scenarios and compare them for the sum of the cooperation gains they produce, over the long-term, for both actors individually or combined.[[1]](#footnote-1)

Cooperation gains for each actor are equal to the utility they reap, based on the definitions of their utility functions. Actors’ utility is calculated as the negative of the Euclidean distance of the particular outcome from the actors’ ideal points, so $U\_{A}^{X}$, utility of actor A from point $X$, is equal to: $U\_{A}^{X}=-d\left(A^{\*};X\right)=-\sqrt{(A\_{1}^{\*}-X\_{1})^{2}+(A\_{2}^{\*}-X\_{2})^{2}}$, where the lower subscripts on terms under the root denote the two dimensions of the political space in Figures 1 and 2. This also means that utilities are always non-positive. Further, as per the utility function description in the body of the article (Section 3), the Challenger faces the one-off costs of the setup or adjustment of the institution (competitive regime creation or regime-shifting), denoted by the component $c\_{f}$ in the utility function $U\_{C}^{X}=-d\left(C^{\*};X\right)-c\_{f}$.

Let us present the two scenarios, corresponding to the logic presented in Sections 3 and 4. In the first scenario, the Challenger decides to engage in CM. The entire interaction starts with this step in which the Challenger and the Defender move from the status quo to the CM outcome. As a result, the actors are in a situation of rivalling rules enshrined in two overlapping institutions (“strategic inconsistency”). As theorized in *Proposition 1*, this should lead to an overall loss of cooperation gains. The process of inter-institutional accommodation starts to unfold then. The process lasts for a certain (randomly selected) number of rounds, after which the actors reach inter-institutional complementarity, as theorized in *Proposition 2*. From then on, and for the rest of the game, both actors reap gains corresponding to this new arrangement. In the simulation the new arrangement is located in the vicinity of or at the Nash bargaining solution; i.e., it tends to be symmetrical and efficient. The sum of the (discounted) payoffs from these two phases, jointly lasting for the 100 rounds of the game, then constitutes the overall gains under this CM scenario. The fixed one-off costs for institutional setup are subtracted from the Challenger’s payoff in this scenario.

Second, in our baseline scenario the Challenger does not move to CM. By implication, the status quo in the focal institution persists for some time. After a random number of rounds, however, a gradual decline of cooperation gains sets in due to the effects of misalignment of the institution and the prevailing distribution of power (see Section 3). This slide of decreasing gains continues until the level corresponding to gains from CM is reached. That is, over time, the actors produce a spiral of non-compliance that is likely to continue until it reaches the level at which they converge towards their respective within-group preferences. This point corresponds to the CM outcome in terms of payoffs, but is characterized by only one institution, the focal institution, becoming weakened by the spiral of non-compliance. This point is reached in the 100th round, where the game ends.

The comparison between the two scenarios outlined here, with and without CM, constitutes the basis for our analysis of the effects of CM on cooperation gains. Although the utilities of actors are always non-negative (utility reflecting the negative of the distance of a particular point from actors’ ideal points), the comparison of the two scenarios produces either negative scores (payoff when the challenger pursues CM is lower than in the baseline scenario) or positive scores (payoff when CM is pursued is higher).

To provide a full picture of all key elements of the game, we also explicitly model the conditioning effects of timing on the overall balance of gains and losses from CM. The logic of our argument in developing *Proposition 1* (Section 3) suggests that if the focal institution is very stable, CM is likely to have stronger negative effects because it represents an inefficient departure from stable cooperation. Similarly, the logic underlying our discussion in Section 4, leading to the formulation of *Proposition 2*, expects that the gains that states will reap from CM depend critically on how long it takes them to realize inter-institutional complementarity after CM has taken place. The longer the states are stuck in the inefficient CM outcome, the stronger the overall negative effect of CM on cooperation gains.

Our approach is to let these factors vary randomly within a meaningful range across a high number of iterations of the simulation.[[2]](#footnote-2) A number of further input parameters needs to be specified in the model. Table A1 summarizes them with their ranges. First, as already mentioned, two parameters of particular interest are the stability of cooperation under the status quo and the duration of inter-institutional accommodation once CM has occurred. Both parameters vary between 0 and 100 rounds. In addition, we also include several other factors, especially the initial location of the status quo, which directly implies the degree of its bias and inefficiency. Another parameter that we randomly vary pertains to actors’ patience. Finally, we let the fixed costs of institutional change vary between one and ten equivalents of per-round payoff from the status quo, i.e. between one time and ten times the equivalent of payoff the challenger would reap from the continued status quo.

**Table A1**: Computer simulation parameters

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Range of values (input drawn randomly from a uniform distribution across the indicated range)** | **Note** |
| Duration of the process of inter-institutional accommodation after the exercise of CM | 0-100 rounds | This duration value also directly determines the length of the new cooperative inter-institutional complementarity, as combined these amount to 100 rounds. |
| Stability of the status quo | 0-100 rounds | This value also determines the length of the institutional decay process under the baseline scenario, as combined these amount to 100 rounds. |
| Location of the status quo | Random values within the colored area in Figures 1 and 2 |  |
| Patience (delta parameter in discounted future gains evaluation) | 0.9-1 | Future gains discounting based on formulas in McCarty and Meirowitz (2007, sec. 3.5)  |
| Location of the Challenger’s ideal point C\* | Located in range between 0 and 10 on the horizontal axis and between 90 and 100 on the vertical axis in Figure 1 |  |
| Location of the Defender’s ideal point D\* | Located in range between 90 and 100 on the horizontal axis and between 0 and 10 on the vertical axis in Figure 1 |  |
| Location of new inter-institutional complementarity outcome | Located in range between 45 and 55 on both axes in Figure 1 |  |
| Transaction costs of institutional change ($c\_{f}$) | 1-10 times the per-round payoff challenger receives under continued status quo |  |

*Note: The space is normalized to size 100x100 where three corners of the square are defined by CM, C\* and D\* (see the upper-right part of the spatial model in Figures 1 and 2). The degree of status quo bias and inefficiency is calculated directly from the status quo location.*

## 2. Assessing the Overall Effects of CM on the Realization of Cooperation Gains

What overall effects does CM have on the realization of cooperation gains under our model? The core results pertinent to our Propositions 1 and 2 are summarized in Figures A1 and A2. Figure A1 shows the sums of absolute values of both actors’ combined payoffs in four different situations, forming the two scenarios we study.[[3]](#footnote-3) The colored boxes capture the gains from the baseline scenario without CM. The empty boxes capture the gains under our scenario of interest where CM is pursued by the Challenger. The results of the first phase of the game are depicted in the left part of the figure. The top-left box shows the joint payoffs under the status quo, with the existing (and functioning) focal institution. The bottom left (empty) box shows the distribution of joint gains reaped by states in the first phase of the CM scenario, i.e. the gains states jointly reaped immediately after the challenger decides to pursue CM. The results closely support Proposition 1, as gains reaped under CM are significantly lower than those obtained under the status quo. The choice to engage in CM is associated with a dramatic drop in joint gains. This corresponds to the insights from Figure 1 in Section 3 where, under CM, states move toward a collectively less efficient outcome than exists under the status quo.

In each simulated game, after a random number of rounds, the actors move to Phase 2. Under the baseline scenario this means there is a gradual decay in gains from the status quo level. The colored bottom-right box shows the distribution of these joint gains. Under the CM scenario, a new inter-institutional complementarity is reached by the actors; and for the remaining number of rounds they reap gains corresponding to this efficient and symmetric outcome. These are summarized in the top-right empty box. Again, the results closely support Proposition 2, where the pursuance of CM, when followed by the re-opening of the deadlocked institutional bargaining and by the new inter-institutional adjustment, provides superior results to pursuing the declining institutional status quo.



**Figure A1**: Gains from different phases of interaction under the alternative scenarios

Figure A2 depicts the overall net balance for both these situations, i.e. the distribution of the net effects of engaging in CM in the two phases as depicted in Figure A1. The figure shows, quite clearly, that, in the first phase, the overall effect of pursuing CM is strongly negative. In the second phase, however, if inter-institutional complementarity is reached, as we argue it may be, the joint gains for both actors sizably outweigh the gains they would reap under the declining status quo. This provides further direct support for Propositions 1 and 2.



**Figure A2**: Net gains from CM, compared to baseline scenario, for both actors combined from different phases of interaction

Can we say anything about the overall balance? Whether the negative or positive effect prevails depends on the actual values of the various input parameters, most prominently on the factors of timing (see below). With the simulation input parameter values drawn from distributions with ranges as summarized in Table A1 above, we find that around 91% of the possible parameter constellations lead to a loss of cooperation gains for at least one actor and 82% for the overall loss for both actors combined (i.e. the lowering of the amount of joint gains available). In roughly 9% of the possible input parameter constellations, both actors gain from CM; in more than 18% of cases, the additional gains reaped by the Challenger from CM outweigh the losses suffered by the Defender. In other words, clearly under a range of circumstances, though fairly limited in this setup, the overall effects of CM can be positive. Later, in Table A2, we show how these ranges of conditions can expand when only certain ranges of the input parameters are considered. Having said that, the specific figures reported in this appendix should only be seen as manifesting that our reasoning formalized in the spatial model in Figures 1 and 2 translates also into dynamic setting, not necessarily as empirically *a priori* credible estimates.

## 3. The Stability of the Status Quo in the Focal Institution and the Duration of Post-CM Bargaining

When formulating Propositions 1 and 2 in the body of the text, we suggest that two time-related parameters condition the positive and negative effects of CM on cooperation gains. These are the stability of the focal institution under the status quo, and the duration of the inter-institutional accommodation that follows CM. We now proceed to evaluate the effects of these two parameters in our simulation setup. The *first* parameter is operationalized as the per cent share (number of rounds out of the total of 100) of the iteration of the game that the cooperating actors spend in the original status quo of the focal institution, before the decline in compliance and institutional decay sets in. The second parameter of theoretical interest is the time that actors spend in the process of seeking inter-institutional complementarity, i.e. in the gradual move away from CM, towards inter-institutional complementarity (see Figure 2 in the body of the article).

Figure A3 depicts evidence pertinent to the two parameters. Figure A) on the left hand focuses on the stability of cooperation under the status quo, plotted on the horizontal axis. On the vertical axis, we show the percent share of the input parameter constellations that produce an overall positive outcome of CM on cooperation gains (full black line) or that produce positive outcome also for both the Defender and the Challenger individually (i.e., is Pareto-improving, dashed red line).

The Figure clearly supports our reasoning. First, as the stability of the focal institution increases, the overall effect of CM grows more negative. In other words, with a highly stable focal institution actors do not gain collectively by pursuing CM; they can continue their cooperation for a very long time, in spite of the inequality in the gains distribution it produces. At the same time, there are specific constellations under which the overall effect is positive. Out of all the constellations, around 30% produce an overall positive effect of CM when the status quo is extremely unstable and institutional decay of the status quo starts immediately. When the status quo is highly stable, as in the right part of Figure A), only around 10% of constellations of input parameters produce an overall positive CM effect. In general, approximately half of this share of parameter constellations satisfies both the Challenger and the Defender, as captured with the dashed line, declining from less than 20% to around 5%.



**Figure A3**: The effects of the stability of the original status quo (A) and of the duration of inter-institutional bargaining (B) on the share of input parameter constellations under which CM creates gains.

Figure B) demonstrates the critical importance of the timing of post-CM bargaining for the overall effects of CM. The construction is analogous to Figure A), except that here the horizontal axis shows the varying duration of the process of inter-institutional accommodation, before inter-institutional complementarity is reached. The higher the values, the longer the inefficient CM point and inter-institutional bargaining with conflicting institutions persists. In line with our reasoning, Figure B) shows that the longer it takes until inter-institutional complementarity is reached, the lower the share of positive-sum constellations. With an extremely short duration of inter-institutional accommodation, i.e. with almost an immediate arrival at successful inter-institutional complementarity, more than 40% of all input parameter constellations produce an overall positive sum of gains from CM. This is very significant, as it shows that high speed inter-institutional bargaining after CM can, under many different constellations, produce an overall positive CM payoff. However, as the bargaining after CM becomes protracted, the share of constellations producing an overall positive gain decreases rapidly to around 10%. Similarly to Figure A), when we only consider the Pareto-improving scenarios that satisfy both the Challenger and the Defender, the share of gains-producing constellations is around half of that, declining from around 20% to around 5%, as the duration of inter-institutional accommodation process prolongs.

These findings confirm the validity of our theoretical reasoning in Propositions 1 and 2 in a dynamic setting, but they also show that whether the positive effect of CM outweigh the negative one depends critically on whether inter-institutional complementarity is achieved quickly once CM has taken place, and whether the original cooperation in the focal institution was stable, in spite of the uneven distribution of gains.

## 4. Robustness Test Regarding Propositions 1 and 2

The outcome variable of interest – the gains states reap from cooperation, and the effects of CM on them – can be measured in two ways. First, we evaluate the overall amount of cooperation gains reaped by both the Defender and the Challenger, combined, in the spatial model.[[4]](#footnote-4) In our aggregate results presentation, as in Figures A1 and A2, we have shown the average joint gains from CM, as compared to the baseline scenario without CM, calculated from a high number of spatial model iterations with varying values of all input parameters. Second, we can evaluate the effects of CM on cooperation by calculating the share of all input parameter constellations that produce an overall joint gains improvement, as compared to the baseline scenario without CM (as in Figure A3). While the two approaches to the measurement of the effects of CM on gains are conceptually distinct, their results co-vary strongly. In Figure A4, we plot the very strong relationship between the two variables aggregated by the durability of the status quo. Pearson correlation coefficient of the two measures is here equal to *r*=0.99 when all input parameter constellations are considered. Aggregation across different input parameters produces similarly strong results, with correlations always well above *r*=0.8.[[5]](#footnote-5)



**Figure A4:** Relationship between the two variants of the dependent variable.

In assessing the overall effects of CM on cooperation gains above (Figures A1 and A2), we estimate that across all input parameter constellations in our model, a joint gains increase is observed in around 18% of cases. We also admit that the specific value varies, depending on what specific ranges of input parameters are considered in the calculation. In Table A2, we show how the share of jointly beneficial CM-constellations varies as we change the values of our input parameters. For each parameter, we report two scores of the outcome variable of interest: one for calculation based on the lower half of the range, and one for the higher half of the range. Note that since the input parameter values are drawn randomly from uniform distributions, the cut point always corresponds also to the distribution median and mean, so the two columns effectively present results for above-average and below-average values of the parameters. The second and the third column show the collective gains surplus (the outcome variable of our interest), while the fourth and the fifth column show gains surplus individually for the Challenger.

**Table A2**: Variation in dependent variable average scores depending on range of input parameters selected

|  |  |  |
| --- | --- | --- |
|  | Share of constellations producing an overall joint benefit | Share of constellations producing a benefit for the Challenger |
| Parameter | Below-average parameter values | Above-average parameter values | Below-average parameter values | Above-average parameter values |
| Duration of inter-institutional accommodation, after the exercise of CM | 24% | 11% | 36% | 25% |
| Stability of the status quo | 22% | 13% | 32% | 28% |
| Patience (parameter δ in discounted future gains evaluation) | 6% | 29% | 14% | 46% |
| Status quo bias | 9% | 24% | 37% | 25% |

As columns 2-3 demonstrate, the share of constellations producing an overall gains surplus changes (in the expected direction), as we move from low to high values of the input parameters. It varies quite widely, between approximately 10% and 25%, strengthening the caveat we raised above that the specific numerical results need to be treated with caution. Highly sizable differences are observed in the parameters that are related to the time dimension of the CM process. In addition, the innate parameter of actors’ patience also proves to play a major role.

As reported in columns 4 and 5, the share of constellations that produce an individual benefit for the Challenger stays, for the most part, between 25% and 45%. It drops to around 14% when the Challenger has relatively low patience, as we discuss further in the last short section of the appendix.

## 5. Actors’ Patience

Since the simulation assumes an iterated interaction, the gains from cooperation are discounted by the factor δ, where payoff P reaped one round away in the future is equal to $Pδ$, payoff two rounds away to $Pδ^{2}$, and so on. The following formula (McCarty and Meirowitz 2007: 47) is used to calculate the discounted payoff $P'$ from individual round payoffs P received over $n$ consecutive periods, as perceived by the actor in time $t=0$.

$$P'=P\frac{1-δ^{n+1}}{1-δ}$$

In the simulation, we let δ vary randomly within the range $δ\in [0.9,1)$. The results presented in the body of the text summarize the calculations across all these values of the δ parameter for the Challenger and the Defender (in each draw of the delta parameter, both the Challenger and the Defender are equally patient). In Figure A5 we show how the relative patience of the actors influences the overall effects of CM on cooperation. Intuitively, the more patient the actors, other things equal, the more likely is CM to bring overall positive results. The effect is non-linear, however, and a much stronger effect on the profitability of CM accrues with high actor patience (the discount parameter $δ$ close to 1). In our simulation setup, the share of constellations producing overall cooperation gains rises quickly from around 5% with parameter $δ=0.9$ to around 60% with parameter $δ=1$. This finding highlights again the critical importance of time for the overall effects on cooperation gains, something that deserves further theoretical and empirical scrutiny.



**Figure A5**: Impact of the Defender’s and the Challenger’s patience on the overall effect of CM on cooperation gains.

## References

Atkinson, A. B. (1999). The Contributions of Amartya Sen to Welfare Economics. *Scandinavian Journal of Economics*, *101*(2), 173–190. https://doi.org/10.1111/1467-9442.00151

Hinich, M. J., & Munger, M. C. (1997). *Analytical Politics*. Cambridge: Cambridge University Press.

McCarty, N., & Meirowitz, A. (2007). *Political Game Theory: An Introduction*. Cambridge: Cambridge University Press.

1. This utilitarian approach of assessing cooperation by looking at the overall gains it produces may be criticized for being agnostic with regard to distributional issues as such (see e.g. Atkinson 1999). Yet we deem it appropriate as a starting point for analyzing state behavior in international politics, where no strong norms for an appropriate distribution of gains and wealth, or for what constitutes equality, exist. [↑](#footnote-ref-1)
2. In total, we have replicated the simulation with 1 million different combinations of the input parameters’ values. [↑](#footnote-ref-2)
3. To allow for a comparison, the gains in this graph are not discounted. Given that the game has 100 rounds, if gains from later phases were discounted the results would be visually close to meaningless, as the phase-two values would be very low. [↑](#footnote-ref-3)
4. In the spatial model construction and analysis, we follow Hinich and Munger (1997). [↑](#footnote-ref-4)
5. With aggregation across some of the input parameters, the relationship between the two versions is strongly nonlinear. In its raw form, whether or not CM produces positive outcomes is in every particular case a binary matter, hence aggregation is needed to allow for the checking of a correlation between the two alternative measures of our dependent variable. [↑](#footnote-ref-5)