

Dear Editor and Referees,

I am yet again grateful for your very constructive feedback on this manuscript. I have ...

What hasn't changed ...

There is one significant change that is not noted below ...

I have once again written a consolidated response. I have grouped your responses together where they are related and responded just once wherever possible, with my responses in italics.

With sincere gratitude,  
Kristy

## 1 Clarify statement of lobby's incentive constraint

### Editor Point 2:

I think the way you write and analyse the lobby's incentive constraint is still not very clear. I would explain things in the following sequence: (a) the relevant level of  $e_b$  in constraint (8) is the best deviation effort, i.e. the optimal effort conditional on inducing a break of the agreement; (b) the best deviation effort is given by  $\max e_{tw}, e^{bar}(\tau^a)$ ; (c) in order for a non-trivial agreement to be possible, there must exist some  $\tau^a < \tau^{tw}$  such that  $e_{tw} < e^{bar}(\tau^a)$ ; (d) to avoid any confusion, I would write constraint (13) evaluated at  $\max e_{tw}, e^{bar}(\tau^a)$ , rather than at  $e^{bar}(\tau^a)$ . Part of the reason I found the current exposition confusing is that your constraint (13) assumes  $e_{tw} < e^{bar}(\tau^a)$ , but this restriction is imposed and discussed only after writing constraint (13).

### Referee 2 Point 3:

I believe (8) only need hold for  $e_b \geq \bar{e}(\tau^a)$ . I believe (8) would fail for  $e_b = \bar{e}(\tau^a) - \varepsilon$ , right? This seems confusing because (8) seems to be stated for any  $e_b$ .

### Referee 2 Point 5:

I found the program on page 18 a little confusing. Equation (11) puts a constraint on  $e_b$ , essentially requiring that  $e_b \geq \bar{e}(\tau^a)$  if I understand. Equation (10) seems to be directed toward a value for  $e_b$  such as  $e_a(\tau^a)$ ; that is used on the equilibrium path. If we were to take the program as a mathematical object, (11) would seem to define the range of  $e_b$  that is to be considered in (10). I don't think that is what is intended. I think it might be easier to define the program

using  $\bar{e}(\tau^a)$  instead of waiting to define that function later.

**Referee 2 Point 6:**

Equations (12) and (13) finally define the IC constraints using  $\bar{e}(\tau^a)$ . In comparison to (7) - (10), (12) replaces  $\tau^R(e_b)$  with  $\tau^b(\bar{e})$ . Is there a difference between the  $\tau^R$  and  $\tau^b$  functions? If these are referring to the same functions, I suggest sticking with the former notation so as to minimize the burden on the reader. A similar remark applies to (13) in relation to (8) - (11).

- I've

## 2 Better explain argument involving lobby's choice of effort level

**Editor Point 3:**

Probably I am missing something, but I find the argument in paragraph 4 of page 19 confusing. You say that in the punishment phase the lobby chooses some effort level weakly higher than  $e_{tw}$ . But consider an extremely high level of effort: how can this be credible? And how is this statement consistent with (7) and (8), where the punishment effort is  $e_{tw}$ ? And don't you need to pin down the exact level of effort in the punishment phase in order to write the incentive constraints?

- I've

## 3 Remove unnecessary commentary

**Referee 1 Point 1:**

The main drawback of the paper, however, is that it is not an easy read and the analysis is not presented well. For example, while presenting the formal model, the author inserts various informal discussions that are very lengthy and unnecessary. Most of these discussions are provided to justify the real-world relevance of the model's assumption. I think the author could greatly improve the presentation of the model by treating it as a purely theoretical model. For example, I don't think that the following paragraph from page 25 adds any insights to the theoretical discussion in section 5:

"A change in  $\delta_L$  might reflect a change in firms' planning horizons, or even their operational horizons although it is not entirely clear in which direction this might work for firms who are facing extinction without sufficient protection. The lobby's patience level might also change with a change in the administrative leadership of the lobby, or as a reduced form for changes in risk aversion in a model with political uncertainty a more risk-averse lobby would effectively weigh the future, uncertain gains less relative to the current, certain cost."

There are various paragraphs similar to this one that can be simply removed from the paper.

- I

## 4 Proofread thoroughly

**Referee 1 Point 2:**

The paper still has various typos and a thorough proofreading is needed.

- I

## 5 Clarify definition of break tariff

**Referee 2 Point 1:**

It seems that that (7) implicitly assumes that  $\tau^R(e_b) > \tau^a$ . Is that right? If so, please be explicit.

- I

## 6 Reconsider labeling of break effort level

**Referee 2 Point 2:**

It is a little awkward to use  $e_b$  to call to mind a "break" when under (7) we have that  $e_b$  is set so that the ML does not break the agreement.

- *I agree that this is awkward. I have tried to come up with some terminology and notation that means "tries to disrupt the agreement but is not necessarily successful. I have not been at all successful. I apologize.*

## 7 Add graph?

### Referee 2 Point 3:

I suggest that the author consider adding a figure with  $\tau$  on the y-axis and  $e_b$  on the x-axis, and with an upward sloping line corresponding to  $\tau^R(e)$ . The author could then depict  $\tau^a$  and  $\tau^{tw}$  in ascending order on the y-axis, and similarly  $e_a(\tau^a)$ ,  $e_{tw}$  and  $\bar{e}(\tau^a)$  in ascending order on the x-axis. Given  $\tau^a$ , we can find  $e_a(\tau^a)$  off of the  $\tau^R(e)$  curve, and similarly for  $\tau^{tw}$  and  $e_{tw}$ . If the distance between  $\bar{e}(\tau^a)$  and  $e_{tw}$  is large in comparison to that between  $e_{tw}$  and  $e_a(\tau^a)$ , then it can be seen that the cost to the lobby of going sufficiently above its ideal point,  $e_{tw}$ , could offset the future benefit of being eliciting  $\tau^R(\bar{e}(\tau^a))$  and then  $\tau^{tw}$  over the punishment phase. I may be mistaken, but I think this is the basic tradeoff that sits at the foundation of the analysis.

- I