We are very grateful for the very constructive feedback and for the opportunity to revise this manuscript for consideration in the *QJPS*. Our responses to the editors and both referees are below in blue.

Best regards,
Kristy

—

Dear Kristy,
We have received and are enclosing two referee reports on your paper, "Inefficient Concessions and Mediation." As you can see, the referees' responses are mixed. R1 is critical of the model and makes several good points. R2 is more favorable, but their prescribed revisions leave a long road between the current draft and the one that they would support for publication in the *Q*. We agree with R2 and see potential in the model, but also believe the revisions required will be significant, and there is still some uncertainty about the outcome when the revision is completed.

If you would like to proceed under these conditions, we would be happy to receive a revised version of the manuscript for continued consideration. But we would like to note that this is a more precarious R&R than usual.

If you choose to take us up on this offer, the editor will review the revised manuscript and we will send it back to reviewer 2. In addition to addressing as many points as possible from R1 and satisfying R2, we would also like you to address a few explicit revisions suggested by the associate editor.

The first and most important point is that both the editor and R2 think this is a paper about trust and cooperation in a collective good framework and is not really a "conflict" model. The manuscript runs into issues with R1 because its pitches the analysis as speaking to the bargaining theory of war, and commitment problems in that context, but doesn't really connect to those models. As R1 points out, given your model you end up with results that are backward (I.e., wanting to fight when delta is low) and there is no real explanation of why this is the case. It seems clear that the answer is because the model and the traditional story of commitment problems and conflict do not line up. What is really happening in this model is that there is an opportunity for collective action (specifically cooperation) in a prisoners' dilemma setting and the players don't know if their partner values the future sufficiently for the long-term benefits of cooperation to trump the short-term incentive to cheat. The manuscript then proposes unilateral concessions to demonstrate trustworthiness and considers how different situations, for example mediation, might make such cooperation easier or more difficult.

We agree that it is useful to frame this model as more broadly about public goods provision, with peace as a prominent example of a public good that can be provided. This broadening is now reflected throughout the revised manuscript.

In particular, the revised introduction is organized around three archetypal cases of inefficient concessions that motivate the model: disarmament, strategic territory, and the export of advanced technology to strategic rivals. This allows us to focus, from the outset, on the application of the model not only to military conflict, but also to a broader set of cases in which distrust undermines the provision of peace and other public goods – e.g., economic growth and technological progress. To underline the broader applicability of the model, we have also changed the language of the model from Peace/War to Cooperation and No Cooperation, as well as Fight to Distrust (p. 11).

With respect to the low-delta result, this emerges because the impact of patience is modeled differently in our set-up than in the bargaining model of war. We have altered the framing of the paper to no longer focus on the bargaining model of war – we have removed these paragraphs, while retaining citations to this literature in a footnote. Hopefully within our new framing this low-delta result no longer appears counter-intuitive. (We have also added a footnote that explicitly contrasts the role of patience in our model and in the bargaining model of war) (p. 4).

To both the associate editor and R2, this is an interesting problem, and is related to both the frequent occurrence of gift giving and unilateral concessions in early modern international relations, but also its relative infrequency today. A puzzle is: if gift giving and concessions were a trust building signal in the past, why was it useful, and why has it become less prevalent? Or maybe you think it still happens often and we don't recognize it as such.

This is an interesting idea regarding the evolution of gift giving over time – our focus is quite contemporary and so this is not something we had thought through. There are quite a lot of differences between the strategic environment facing states in early modern international relations than in the modern period on which we focus, and a number of these differences could plausibly shape the relative frequency of gift giving between the two periods.

One difference of relevance to our model is a likely difference in *p,* the proportion of states that are high type. A high type is a state that values public goods sufficiently to make them willing to invest in cooperation to increase the likelihood of long-run consumption of peace or a similar public good. It seems reasonable that in the modern period, with a significant number of democratic states, as well as a number of authoritarian states that rely substantially on performance-based legitimacy to remain in power, that the proportion of high-types is higher today than it was in the early modern period.

If states are, on average, more patient, then Corollary 1 tells us that a separating equilibrium with no concessions is preferable to one with concessions. If democratic states are also more patient in that they place more value on the welfare of future generations or their leaders have longer time horizons, then the *No Concessions Separating Equilibrium* is available to a larger proportion of state actors.

With this approach and focus a few things become clearer. First, "commitment" is really far-sighted preferences that allow cooperation to emerge. In the interesting parameter space, uncertainty results in missed opportunities for cooperation. Concessions and mediation then emerge as natural solutions to this problem, and then the analysis unfolds naturally.

Commitment as a far-sighted preference (p. 3) is a helpful way to explain how commitment works in our modeling set up. Thank you.

Second, there are a number of interesting points of theory that are left unexplored that would add to the value of the analysis. Some that could be addressed are:

1. Are there semi-separating equilibria with mixing by the low resolve type? If so, when, and if not why.

   Semi-separating equilibria with the low types mixing would not allow for the information transmission that is required for a Concessions Separating Equilibrium. Since a low type might give a concession in this equilibrium, high types cannot use the provision of a concession to learn anything new. Thus, the low type giving concessions in this situation would not create the required separation.

   It would be interesting if there were an equilibrium where the low types do not separate but both (some of) the low types and the high types give a concession. Here, it would seem that the high types are tricked into cooperating with a partner who will not cooperate, but a player cannot be 'tricked' in this way under an equilibrium concept. We can see this because all the behaviors in the repeated game would be the same in this proposed semi-separating equilibrium as in the *No Concessions Separating Equilibrium*. The only difference would be that, in equilibrium, the high type always gives a concession but only receives a concession from other high types and the low types who give a concession. But high types would be better off not giving the concession while receiving it from some low types; this profitable deviation destroys the possibility of such an equilibrium.

   We've added the following footnote to the paragraph immediately after Lemma 1 (p. 15):

   > The transmission of information about each country's type is, by definition, essential for the existence of a *concessions separating equilibrium*. Thus, there cannot exist a *concessions semi-separating equilibrium* where the low type gives a concession with strictly positive probability; the concessions would not transmit the information required to build trust. If no information is transmitted, then the high type has no incentive itself to give a concession. Therefore, there cannot be a semi-separating equilibrium with the low type mixing between strictly positive and zero concessions and the players separating in the repeated game as in the *no concessions separating equilibrium*.

   Interestingly, there *can be* a semi-separating equilibrium where the high type mixes. In this equilibrium, (1) the concession is smaller; (2) a low-type benefits less from receiving concessions from high types in the fully-separating CSE (because, in expectation, fewer high types give concessions); (3) and the high type is trading off the smaller expected concession with less of a possibility of learning about a match with a high type that would allow peace. However, in order for the indifference condition to hold, we need $\delta = \frac{W}{T+D}$, which is the (weak) lower bound on the level of patience for the *concessions separating equilibrium*.

Because this semi-separating equilibrium only occurs for a knife-edge case, we don't include a full treatment in the text. We have, instead, added the following footnote to the second paragraph after Theorem 2 (p. 16):

> When $\delta = \frac{W}{T+D}$, there is also a semi-separating equilibrium where only some high types give a concession. Here, high types give a smaller expected concession and trade off a lower probability of achieving peace. We will thus focus on the pure *concessions separating equilibrium* in the analysis below because it provides the best chance for peace and is possible over a much larger portion of the parameter space.

2. What is the relationship between the size of the concession, the likelihood of cooperative types, and patterns of gift giving?

   This is a productive direction to be pushed. We have added to the discussion of comparative statics near the end of Section 4.1 (p. 21). The complete statement now reads:

   > The other variable that influences the patience threshold in the case of future value is p, the proportion of cooperative types in the population. Although the gift in the no-future-material-value case is an increasing function of p, it is a linear function of p just like the variables T, W, and D. Thus the cost of paying the gift and the direct benefits in the repeated game increase at the same rate so that the likelihood of cooperation does not vary in p when gifts have no future material value. In the future-material-value case, the size of the gift required to separate the high types from the low types increases more than linearly in p and thus faster than the other terms. Therefore the patience threshold in the future-material-value case increases as p increases. That is, the cost of giving the gift weighs more heavily in the decision to cooperate relative to the material costs and benefits, requiring states to be more patient to separate and thus reducing the likelihood of cooperative types.

3. Does the modern period have fewer concessions because they would have to be larger to separate?

   As noted above, we think that modern and early modern periods may differ with respect to *p,* the probability with which a state is a high type. As described above, when p is higher, larger gifts are necessary to separate between high and low types. This could plausibly be related to changes in patterns of gift giving between the modern and early modern periods. However, we are somewhat hesitant to speculate in this direction, given the number of other strategic dimensions that also vary between the periods.

4. Can you be trapped in a concession equilibrium that is uninformative?

An uninformative equilibrium is ruled out by definition of the *concessions separating equilibrium*. In this equilibrium, the magnitude of the concession is the smallest amount that will deter the impatient type from mimicking the patient type. Countries who give the concession are immediately revealed to be high types, and those who don't give the concession are immediately revealed to be low types.

The point is clarified a bit by thinking about what happens if concessions below the separating threshold are given. In this case, any low type will give the concession just like high types; the negotiating partner will know this and thus the concession will not be informative. There will be no benefit to giving the concession because the negotiating partner does not learn anything and thus does not change its behavior. Thus there is no incentive to give a concession that is not informative.

After "low types do not give concessions" on page 15, we've added: "because the concession level is set as the smallest amount that will deter the impatient type from mimicking the patient type. In contrast, high types always give a concession in period 0. Thus, the patience level of both countries is fully revealed in equilibrium."

5. Is there anything interesting happening with the dynamics? What if you imposed something like weak renegotiation-proofness on the punishment strategies, do gifts get you more cooperation?

   Imposing a renegotiation-proofness concept would not change any of the outcomes. This is because the low type will never cooperate, so there is no better outcome to which the high type would want to switch away from the punishment in the repeated game.

There are a couple other revisions of presentation that are needed as well. For example, the mediator section needs more explanation. Sometimes it reads like a mechanism design problem, describing what is possible in a PBE. Other times it reads like actual mediation as an institution. This makes it hard for the reader to understand what you are trying to do here and insert their own interpretation. Sometimes the result is favorable (R2), sometimes it is not (R1).

Thank you for raising this. We have re-written the section to explain that we are modeling the mediator as solving a mechanism design problem (p. 24) and putting the focus squarely on the mediator who *uses* the mechanism to choose the concession levels.

In terms of placing this piece in the literature, the manuscript needs to make its relationship to two existing sets of research clearer, the burned money signaling literature and work like Farrel and Gibbons paper on cheap talk and bargaining. With the new framing related to cooperation and trust, you should also flesh out the connection to existing work in that area (some of which you cite already like Kydd and others.). Finally, there is some work you cite on repeated games, cooperation, and incomplete information about delta. Being clearer about your contribution there would be helpful. We understand that we are asking for an extensive revision and that the outcome remains uncertain, but we do believe the manuscript has significant potential, should you decide to take up our offer.

We have revised the framing of the paper significantly, especially related to applications of the model to the provision of public goods outside the domain of peace/conflict. We have expanded footnote 11 (p. 10) to be clearer about the relationship between our contribution and that of Maor and Solan (2015) in the repeated games and incomplete information literature, and have revised our engagement with the literature more broadly as well, limiting our discussion of the bargaining model of war and expanding elsewhere.

Please let us know if you intend to make the required revisions and, if so, when you might be able to return a revised manuscript. We will read the revision and your memo outlining your reactions to the reviews and any changes you have made, and then we will decide whether to go back to a second round of external review. We have attached a copy of the QJPS Style Guidelines. Compliance at this stage would be helpful should we ultimately decide to accept the paper.

We greatly appreciate the work that has gone into this – two thoughtful and detailed external reviews and this editor's memo. We have made significant revisions throughout the paper and we hope the editor and reviewers find the revised paper much improved.

Thank you again for considering the QJPS as an outlet for your best work, and we hope to see a revision soon.
Best wishes,
Kris, Anthony and Stephane

Reviewer 1's comments
This paper contributes to the literature on war due to commitment problems, and finds that if a mediator can be used to reduce the commitment problem then it helps... Not surprising at all. The increased salience of the commitment problem when a transfer concession is made is studied also in Jackson and Morelli (2007), and in that context nobody has studied whether a manipulative mediator who can enforce outcomes could decrease the concern. But again a positive answer to this question in any model is hardly surprising.

It is useful to be pushed on this – we hope that the revised framing of the paper helps make our contribution more clear. We don't view the headline result about the usefulness of a mediator to be particularly surprising. We view the contribution in this paper in this regard to be showing that the result that mediators can lead to greater efficiency is broader than previously established (more on this below).

In the revised introduction we state more directly that the results of our model reveal a type of mediation that "can achieve peace or the provision of other public goods (1) in situations where bilateral concessions cannot, and (2) with concessions that are more efficient when bilateral concessions must be inefficient to achieve peace" (p. 9). In a context in which there is significant debate about the utility of mediators, we believe the additional rationale for mediation that we introduce here is substantively important and points to real-world applications of manipulative mediation that can facilitate the provision of peace and other valuable public goods.

The second concern about the paper is that the specific assumption that the authors use in order to link the role of the mediator to the commitment problem is the following: they assume that the asymmetric information is about the ability to commit of players, i.e., there is some prob that the opponent is high or low type in terms of commitment ability. In this way, if a direct revelation mechanism imposed by the mediator reveals that the players are committed types, then the endogenously informed mediator can enforce the outcome at no cost. In other words, the claim is that a mediator (a manipulative one who can enforce outcomes, in contrast with the pure information and negotiation design role usually invoked when enforcement by third parties is not possible) can facilitate more efficient concessions by removing uncertainty about the ability of the parties to commit to peace.

We believe that modeling uncertainty over the ability to commit to cooperation is a valuable contribution to the literature. This kind of uncertainty is realistic and understudied. If we understand the reviewer’s comments correctly, the concern is the reduced-form way in which we’ve modeled uncertainty over commitment ability. Because this paper focuses on how uncertainty about the ability to commit affects cooperation and concession outcomes, we have chosen to put the modeling focus on those outcomes. We think this generality is an asset of the model, in the sense that the ability/inability to commit could derive from a number of underlying models.

Third, beside the conceptually problematic assumption on the possibility that Nature chooses a commitment ability type, rather than studying commitment as a decision problem in a dynamic game, the modeling choice is problematic for an additional reason: The type is modeled using high and low discount factor, assuming that the low discount factor player is the one always going for conflict. This contrasts with a large literature that actually sees patience as a source of conflict risk. At the very least such a literature should be addressed, and explain why an infinite horizon repeated prisoner dilemma game with private information discount factors is a more appropriate model than those where patience works the other way because of a richer action space and stage game.

This is a valid concern. As we previewed in our response to the editor earlier in the memo, the impact of patience is modeled differently in our set-up than in the bargaining model of war. We have altered the framing of the paper to no longer focus on the bargaining model of war – we have removed these paragraphs, while retaining citations to this literature in a footnote (p. 4). Hopefully within our new framing this low-delta result no longer appears counter-intuitive.

We have also added a footnote that explicitly contrasts the role of patience in our model and in the bargaining theory of war. The footnote reads, “This is in contrast to the bargaining theory of war, where a high(low) type has strong(weak) military capability that then interacts with a symmetric, known level of δ. This accounts for the different impact of δ in the two models” (p. 10).

A fourth concern relates to the connection between the model and the normative evaluation: The timing of the model is described as follows: Stage 0: concession g_i is chosen; stage 1: infinite horizon prisoner dilemma game with private information on the discount factors. The concession separating equilibrium where high types make concessions and hence if at stage 0 both make concessions then they cooperate for sure is the potentially interesting equilibrium, subject to the caveats and doubts mentioned above. Indeed, Theorem 4 is the main result before considering the role of the manipulating mediator: there are intermediate values of the high type's patience for which the efficient no concession separating equilibrium does not exists but the concession separating equilibrium exists, especially when concessions have future value. It is precisely in this range of parameters that an enforcing mediator can further increase welfare of high types. But is this a good measure of welfare? "We will take the measure of social welfare, and thus the determinant of the optimal equilibrium, to be the sum of participating high types' expected utilities." This assumption is not motivated, and contrasts with what people usually do when applying a signaling model like that of Spence.

We have revised the text to make more explicit our motivation for definiting welfare in this way. We now write, "We take achieving peace, or more broadly, achieving the provision of public goods, as the normative goal with respect to advancing social welfare. Only high types are capable of making peace/providing public goods, and thus we will take as the measure of social welfare to be the sum of the participating high types' utilities" (p. 12).

Minor comments:

1. Tone down in introduction the statements on impossible peace with efficient concessions, better say that sometimes an efficient concession can exacerbate the commitment problem the government faces. Making commitments to public job sharing like for IRA or in terms of political access like in the case of FARQ are necessary conditions, whereas making one type or another of weapon disarmament, efficient or inefficient, may be second order.

We have rewritten the introduction considerably and have toned this down as suggested.

2. I find the discussion of the literature on inefficient gifts a stretch. It is fine to underline the connection with the modeling of Camerer, but the analogy with gifts ends there, and hence the rest of the lit discussion in my view is a useless distraction.

We worked to shorten this section by, among other things, cutting our discussion of Prendergast and Stole (2001) (p.6).

3. In the theoretical literature Horner et al 2015 and Meirowitz et al 2019, on restud and jpe respectively, are missing.

Thank you. We have added these. The Meirowitz citation has proved particularly important (see final paragraph of the introduction, p. 9).

4. Footnote 11 uses language inconsistent with standard theory of repeated games.

We have revised this footnote (footnote 12, p. 10) to improve clarity. It now reads: "For a game-theoretic analysis of the repeated Prisoner's Dilemma with uncertainty over discount factors, focused on studying players' belief structures, see Maor and Solan (2015). Our model differs in that it adds an initial period, before the repeated Prisoner's Dilemma begins, in which players make costly concessions to each other. We will see that these concessions enable signaling between the players that render the analysis of their beliefs much simpler than in Maor and Solan (2015).”

5. Interpreting Cyprus, Sri Lanka, Israeli conflicts, all as examples of the pooling equilibrium in which the two players do not give concessions and they fight forever in stage 1 is a stretch, because it is not clear at all that the main feature of such cases compared to others is the uncertainty about commitment types.

In each of these cases, we think that one reasonable interpretation of the historical evidence is that there were concessions that were plausible as means to resolve the conflicts in their early stages, but that these concessions were dangerous to make because they could have been used against the conceding party in future rounds. Each of these cases is complex, and certainly scholars – including those whose knowledge of each case is deeper than our own – might argue that uncertainty about type was not the main feature that prevented these concessions from being made. In our view, we think uncertainty about type likely played an important role in each and that these examples remain useful to the reader.

While we have retained reference to these cases in the current draft (p. 13), if the reviewer feels strongly on this point, we are happy to remove these two paragraphs before publication – the section can stand without them.

Reviewer 2’s comments

Overall Comments

The manuscript examines a two-sided signalling game to highlight the value of concessions and mediators in overcoming the incentives for conflict.

The paper begins with an incomplete-information version of a prisoner’s dilemma game where actors choose either to not engage in trust (cooperate) or to fight the other player (defect). Each actor knows their own discount rate, but not the discount rate of their opponent; conditions are established to identify when, despite low-types always defecting due to their lower discount rate, high types would be willing to cooperate in hopes of matching with another high type and achieving the cooperate-cooperate stream of payoffs.

The paper then adds the possibility of allowing states the option of offering a concession/gift. Conditions for a fully-separating, costly-signaling equilibrium are identified where high types will use the initial concession to separate from low-types. Under some parameters, the equilibrium where high-types give gifts can be better for high-types; essentially, the gift here allows high-types to avoid a “suckers-payoff” from entering into the P-D game and cooperating when the low-type defects. While the gift-giving itself can be a kind of suckers-payoff (because only high-types will give a gift, if a high-type meets a low-type, the high type will give and get nothing in return), it is worthwhile for the high-type because (a) it gives the information of how the repeated P-D game will go and (b) the concession can be lower cost than trusting while the opponent attacks.

The paper then examines versions of the game above where concessions can make the receiver more productive when fighting. This can undermine the existence of the trusting-equilibria. Alternatively, another version of the game is examined where the giver can disable some portion of the concession; the paper finds under some conditions, the optimal equilibrium relies on these kinds of inefficient concessions.

Finally, the paper takes a mechanism-design approach, treating mechanism as a mediator. The paper finds that using the mediator can be most efficient. Essentially, all previous equilibria had some kind of suckers-payoff, where trusting types either (a) trusted while the opponent fought or (b) offered a concession while the opponent gave nothing in return. This is all unappealing to the high-types. By using the mediator, the suckers-payoffs can be avoided; mediators can exchange gifts if both types report they are high, allowing a high-type that meets a low-type to keep their gift while gaining information on their opponent's type.

Because this paper is not seeking to explain empirical results or a puzzling instance of observed state behavior, its merits lie in the novelty and value of the formal results.

It is our view (the reviewer may not agree with us here and that is reasonable) that this paper does seek to explain puzzling behavior: the giving of concessions that are inefficient. We hope there is novelty and value in the formal results as well.

On one hand, as it stands, the results are somewhat flat. One of the key results—that offering a concession can lead to better outcomes for the high-types—is something we'd expect from the costly signaling literature.

We view this not as a key result of the paper, but as a benchmark against which to compare the heretofore unexamined case in which concessions have future material value that can be used against the giver. However, we take the larger point and have addressed it as detailed below.

Another key result—that a "Myerson mediator" can lead to greater efficiency—is something that's been discussed in other conflict papers (see Meirowitz et al, which is cited, but should be discussed more).

Thank you for pointing this out. We have expanded the discussion of Meirowitz et al. 2019 at the end of the introduction (p. 9) and now compare and contrast our contribution more directly. We view the contribution in this paper as showing that the result that mediators can lead to greater efficiency is broader than previously established. That is, the fact that mediation can help could be for a quite different reason (here, it helps remedy incomplete information about the level of commitment to peace instead of the level of militarization) and that it can also improve the efficiency of concessions.

A selling point of the paper could be the examination of what happens when concessions add material value, but currently I have some issues with how this is currently modeled. On the other hand, I think improvements can be made, and some work could be done that better grounds the model in an alternate empirical framing (rather than conflict, considering collective action problems). Also, it may be possible that I am underselling the existing results: I'd be open to hearing more from the authors on this point. Finally: if the authors could derive any additional interesting theoretical results from this model, I would be very open to that.

Thank you for this. We do view the main contribution to be those surrounding material value. We have addressed the specific concerns about how this is modeled (detailed below).

As suggested, we have pushed the paper beyond the conflict context to also address its application to the provision of public goods, such as economic growth and technological progress, in the context of geostrategic competition. We address your more specific comments below, but that is a fairly large-scale set of revisions we made that are reflected throughout the paper.

We have added what we believe are valuable comparative statics and addressed some interesting theoretical points, which are detailed above in the comments to the associate editor.

Concessions With Material Value
In the model where concessions lack material value, should actor 1 fight and actor 2 trust, actors 1 and 2 receive per-period payoffs of $(T + W, -D)$.

In the model where concessions have material value and both actors set $\alpha i = 0$ (implying that the proportion of the received concession that goes to the military is 1), should actor 1 fight and actor 2 trust, actors 1 and 2 receive per-period payoffs of $(T + W + W\ g2, -D - Dg2)$.

First, when introducing material values, the gifts are scaled by multiplying the size of the gift (gi) by some existing parameter (like W or −D). This introduces non-zero sum changes as g2 changes. As mentioned above, a 1 unit increase in g2 gives actor 1 an additional W and actor 2 a deduction of -D, thereby assuming a knife edge case where the value of transfers is linearly related to the existing parameters in the PD game. These are both big assumptions that are not described or justified adequately. If these can be justified, they should. If not, this really should be generalized.

We have included the following in a new footnote in Section 4.1: "The results are qualitatively unchanged if the gifts are not scaled by the variables that represent the impacts of cooperation and non-cooperation, if they are scaled by these variables linearly but with a proportion other than 1, or if they are scaled by many plausible non-linear functions of T, W and D. The results are a bit simpler if there is no scaling, but we believe it is more realistic to assume that the future benefits and costs of the gifts are proportional to the direct benefits and costs of cooperation/non-cooperation. For instance, a country who could impose more damage without the future value of the concessions is likely to be able to better leverage a concession than a country who can impose less damage" (p. 19).

It may be useful to outline the intuition here. Specifically, the results are robust to alternative formulations $(T + W + f(W)\ g2, -D - f(D)g2\ )$ where, for example, $f(x) = \sqrt{x}$ or $f(x) = x^2$ or any linear function of x, including $f(x) = 0x = 0$ (although we do not view increasing returns to be intuitively plausible). For whatever function is chosen, the gift changes from

$g = \frac{(1-\delta_l)p(D+T)}{(1-\delta_l)(1-pT)+(1-p)\,D}$ to $g_f = \frac{(1-\delta_l)p(D+T)}{(1-\delta_l)(1-pf(T))+(1-p)\,f(D)}$ and, similarly the patience threshold changes from $\delta = \frac{p(W-T-D)}{g} + p(W - T) + (1 - p)D + 1$ to $\delta_f = \frac{p(W-T-D)}{g_f} + p(f(W) - f(T)) + (1 - p)f(D) + 1$. The parameter ranges where future material value destroys the possibility of cooperation change, but the qualitative result does not. The results on the effects of mediation are similarly unchanged in a qualitative sense.

Second, the payoffs for the model without material values and payoffs for the model with material values aren't normalized in some clear way (at least from what I could tell?). I think this undercuts Corollary 2: it could be that the material values (who they help and how) drive equilibria existence in interesting ways, but it could also just be that the changing magnitudes of all the payoffs are driving this result. Put an- other way, across these two models, the magnitudes of the payoffs just seem really different. Could different formulations of how concessions help/hurt actors still generate these results?

This is an important point; thank you for bringing it up. The changing magnitudes are directly related to the core idea: that payoffs look *very* different if gifts have future material value. That is, we view this as a design feature and not a design flaw. While it's hard to say there is *no* other formulation for concessions that would deliver results that are qualitatively different, we have confirmed that several different formulations that seem intuitively plausible (see immediately-preceding response, for instance) do not disrupt the results. Analogous results will hold as long as the future benefits of material value from cooperation and the future costs of material value in the case of non-cooperation are treated similarly enough. An example that could disrupt our results: if one believes that the public-good nature of the benefits is long-lived, but the costs are very short lived (or non-existent), then there are parameters under which peace might be achievable for less patient types when there is future material value. That is, the costs must be sufficiently important in order for future material value to disrupt the ability of concessions to build trust.

Is this Conflict/War?

The authors spend a lot of time motivating this paper within the framework of formal models of conflict and mediation. However, this model is quite different from how conflict is commonly modeled; either in a bargaining framework (Fearon 1995), or in a deterrence framework (Fearon 1997, Baliga et al 2020) , or in a deterrence-hawk-dove setting (Baliga and Sjöström, 2020). I think the authors could possibly claim that conflict is analogous to a P-D game, but this is challenging given it historically has not been done that way. Rather, this model fits naturally with any kind of collective action problem (to name a few: climate change, free trade, over-fishing, use of cyberespionage, CFC production, etc). It's not that the authors are trying to fit a square peg into a round hole (the model could possibly apply to conflict), but less attention should be paid to conflict and more attention should be paid to alternate collective action issues that seem a more natural fit.

Thank you. This has proven to be a particularly fruitful suggestion for us and we have reframed the paper to be broader and to incorporate application to a fairly broad array of public goods, of which peace is one. We think that peace remains a powerful motivating case here, but we think the broadening strengthens the paper considerably.

As one proposed restructuring: currently, conflict examples and discussions of conflict are scattered throughout the paper. In most places, this isn't needed (like giving an example of a No Concessions Separating Equilibria—this isn't value added). I would much prefer a consolidated "empirical examples" section or subsection where conflict and other collective action problems are discussed in a condensed manner.

We have struck the specific example you note. We did not create a new empirical examples section, but in our revision of the introduction, we have introduced a more substantial set of three motivating examples, which hopefully provides some empirical grounding for readers before they head into the model.

Thinking about Gifts

One interesting aspect of this paper is how it treats gifts. In bargaining models, an offer is made from actor 1 to actor 2, and where actor 2 has the choice to reject the offer and go to war, or accept the offer and not go to war. In a world without credible commitment, this perspective is difficult to swallow: even if Ukraine gifted Russia the Donbass, there is nothing to say Russia would not keep attacking Ukraine. How gifts are modeled here presents an alternate take on concessions that is outside of the standard bargaining framework and seems quite useful. Note that this isn't the first time this has been done—see Gieczewski's "Evolving Wars of Attrition" as another example. But in short, the way concessions are modeled positions the paper in a different vein from a lot of the bargaining literature, which is a virtue of the paper that should be highlighted.

We have highlighted this aspect of the contribution in the introduction with the following statement:

"Another contribution is that the way we model concessions does not require countries to make an agreement that requires credible commitment, as is common in bargaining models. Instead, countries willingly provide a concession if they expect the signal it sends to provide a benefit in terms of future cooperation and public good provision." We footnote this with: "When mediation is required to achieve peace in the model, it relies upon a manipulative mediator to enforce the required level of concessions."

Minor Points

The paragraph at the top of page 13 is written in a confusing way. It sounds like you will keep on redefining what δh is rather than taking it as given and defining what parameters are needed for high types to "accept."

We have clarified this language in this paragraph. In short, we *do* redefine the patience threshold between low and high types for each equilibrium.

Around page 16, as it is written in the paper, is not clear as to how the gift g1 factors into 2's utility function outside of the PD game (if it does at all).

Thank you for pointing this out. We have clarified the explanation as follows:

> If both countries play Distrust, Country 1 receives not only the immediate value of the concession and then W − D in each period. Its welfare now has two additional terms. First, an additional benefit term in each period of the repeated game $W(1-\alpha_1)g_2$, which accounts for the value of the received concession as well as how much Country 1 invested in the military. Second, Country 1 also incurs extra damages $D(1-\alpha_2)g_1$ in each period of the repeated game that are proportional to the size of the concession it gave and how much of that concession was invested in the military by Country 2. The other payoffs are modified analogously (p. 19).

Footnote 12 seems important: this should be included in the Appendix.

We have added the calculations for the comparative statics to the end of Appendix 7.4.