An application of automated mediation to international climate treaty negotiation

P. Ding[1], M. D. Gerst[1], G. Bang[2], M. E. Borsuk[1,*]

[1] Thayer School of Engineering, Dartmouth College, Hanover, NH USA

[2] CICERO, Oslo, Norway

* Corresponding Author

ABSTRACT

A generic framework for automated mediation of multi-attribute negotiation has recently been reported (Lai and Sycara 2009). This framework seeks to shorten time to agreement, achieve agreements that are closer to Pareto optimality, and remain tractable in situations involving multiple issues, incomplete information, and dynamic reservation utility. These objectives are all relevant to international climate treaty negotiation. Therefore, in this paper, we describe how this mediation framework can be applied to the climate policy setting and articulate a necessary extension to allow for more than two negotiating parties. We then demonstrate application of the framework, employing some simple economic and procedural assumptions. This example shows that automated mediation can add value to the negotiation process without placing an undue mental or computational burden on negotiators. Part of this value comes simply from encouraging negotiators to be explicit about their assumptions and preferences, even if most of this information is not shared with their opponents.

INTRODUCTION

From the 1^{st} Conference of Parties (COP) held in Berlin in 1995 to the 18th COP held in Doha in 2012, representatives of nations have been meeting annually with the goal of reaching agreement on a plan to reduce emissions of greenhouse gases (Gupta 2010). While some effort has been put into attempts to find a burden-sharing scheme that is considered to be fair by all participating nations (Ringius et al. 2002), there is evidence that parties invoke fairness criteria primarily to serve their own material interests (Lange et al. 2010). Thus, the international climate conferences are fundamentally a negotiation over how to allocate the burden of reducing greenhouse gas emissions in a way that each party believes to be of greatest domestic benefit.

More generally, negotiation is a process in which parties with conflicting interests seek to reach an agreement that makes each party better off than they would be without agreement (Nash 1951). While situations of two parties negotiating over a single issue with full knowledge of each other's preferences have been solved by axiomatic methods, including the Pareto-optimal and Nash bargaining solutions, the process by which parties can arrive at such solutions is not so clear (Endriss 2006). It is especially difficult for parties to find an optimal solution when they are unwilling to share information on their preferences.

While the majority of research on negotiation has focused on the two-party, single-issue setting, there has been some progress on two-agent, multi-issue (Fatima et al. 2006) and multi-agent, single-issue settings (Binmore et al. 1986) as well. However, much of this work has continued to assume either complete or probabilistic knowledge of the preferences of opponents (Nash 1950, Harsanyi and Selten 1972, Chatterjee and Samuelson 1983, Lin et al. 2008) and linear utility functions for describing such preferences. In general, and specifically in the climate policy context, negotiations typically involve multiple parties attempting to resolve multiple issues for which they have private, nonlinear preference structures (i.e., utility functions). Further, these parties often do not have a complete understanding of their own utility functions, let alone those of their opponents, over the full range of issues being negotiated. In the real-world, this situation can lead to either a lack of eventual agreement between the parties, or to an agreement that "leaves money on the table" in the sense of overlooking opportunities for win-win improvements.

Intelligent mediation can help negotiation processes reach earlier and better agreements (Stenlo 1972), and mediation by a party internal or external to the COP has been suggested as one strategy for enabling progress (Blok et al. 2005). A mediator could use technical knowledge and preference elicitation, for example, to reveal joint interests or areas of compromise (Sjo'stedt and Penetrante). This information could then be used to construct new proposals or modify existing ones to have a better chance of mutual acceptance.

Unfortunately, the involvement of a mediator in the COP process faces a number of obstacles. It is nearly impossible to expect that any person, especially one who simultaneously represents the interests of a particular nation, could be free from suspicion of bias. Even if a person with impeccable objectivity were identified, parties would still likely be uncomfortable with sharing private information on economic projections and policy preferences. Further, a human mediator would need to be extremely skilled mathematically to translate projections and preferences regarding policy outcomes back into the actions, in the form of greenhouse gas emissions reductions, required by all parties.

Computer-assisted, or automated, mediation (Jennings et al. 2001, Kraus 2001) could play a useful role in the international climate negotiation process. A computer algorithm that queries the preferences of parties, uses these to find win-win opportunities, and makes the calculations necessary to infer the implications for emissions reductions, all while protecting privately-held information and maintaining impartiality, would be of high potential value. Lai and Sycara (2009) recently articulated a generic framework for automated mediation of negotiation involving multiple issues. This framework uses a tractable, query-based algorithm to handle the common situation in which parties do not have fully-articulated utility functions. The mediation procedure is designed to shorten the time to agreement and yield agreements that are closer to Pareto optimality. The focus of Lai and Sycara (2009) is on the two-party setting. Therefore, after reviewing their framework, we describe how it can be extended to allow for more than two negotiating parties. Employing some simple economic and procedural assumptions, we then show how this extended framework can be applied to international climate treaty negotiation. Finally, we discuss the limitations of our approach and opportunities for improvement.

METHODS

Negotiation Setting

We consider N self-interested parties who negotiate over M continuous attributes in a maximum of T rounds. The range of each attribute $j \in \{1, 2, ..., M\}$ can be mapped onto a normalized range of $\Omega_j = [0,1]$ with 0 and 1 corresponding to extreme values of the attributes. Thus, the negotiation domain can be denoted by the unit hypercube $\Omega = \Omega_1 \times \Omega_2 \times ... \times \Omega_M = [0,1]^M \subset \Re^M$, and any point $\boldsymbol{x}$ within this hypercube is referred to as a (multi-attribute) *offer*. We assume that each party $i \in \{1,2, \dots, N\}$ has preferences for offers that conform to an ordinal utility function (not necessarily linear in the attributes) that is not fully articulated, but can be used to compare any two sets of attributes, or *offers*. Further, this utility function is private information, meaning that parties cannot compute the utility of an offer to any other party. We normalize the utility range of each party to [0, 1] with the lower and upper bounds representing the worst/best possible offers in the multi-attribute negotiation space. Thus, the utility function mathematically maps the negotiation space onto the [0,1] range according to order of preference.

It is important to note that, with preferences described by non-linear utility functions, the utility of an offer is not simply the sum of utilities of the individual attributes. Therefore, the parties must negotiate offers as multi-attribute packages, rather than issue by issue. While issue-by-issue negotiations may be more convenient to analyze mathematically, packaged offers allow parties to make trade-offs for different attributes, thus providing opportunities for win-win modifications of existing offers (Zheng et al.). Identifying such opportunities is one way in which a mediator can add value to the negotiations.

When parties negotiate, they must have a criterion to decide whether they will accept any proposed offer. We assume that an offer is acceptable if it is determined to be preferable to that party's best alternative to a negotiated agreement (BATNA) (Raiffa 1982). In the metric of utility, this means that a party is assumed to accept any offer that exceeds the utility of its BATNA. This utility is referred to as its *reservation utility*, $ru_i(t)$, and party i is therefore assumed to accept offer x at time t, if and only if $U_i(x) \geq ru_i(t)$. The set of offers acceptable to each party i is strictly convex, and therefore the *zone of agreement*, Z, defined as the common intersection of the acceptable offers of all parties, is also convex. For a solution to exist to a particular negotiation problem defined in this way, the set Z has to be non-empty, and any member of this set is referred to as a *satisficing solution* (Zheng et al.).

We assume that each party's reservation utility is private information and declines with time. This decline can occur either as a concession to facilitate eventual agreement (by expanding the size of Z) or because the BATNA actually deteriorates over time. This latter situation certainly applies to the climate context, in which harmful greenhouse gases continue to accumulate in the atmosphere until emissions reductions are agreed-upon and implemented.

We assume that negotiations proceed with parties sequentially proposing offers in a pre-specified order to which opposing parties have the opportunity to respond (by either accepting or rejecting). This is a multi-party generalization of Rubinstein's alternating-offer protocol (Rubinstein 1982, Zheng et al.). While our proposed mediation procedure is not specific to any particular proposal mechanism, we adopt a multi-party version of Lai and Sycara's (2009) two-party "linear proposing" heuristic, which can be employed when parties do not know their opponents' utility functions. After a party makes a proposal according to this procedure, we employ a nonbiased mediator – which can be implemented as an autonomous computer algorithm – to query the other negotiating parties and find a point that is mutually better than the standing offer. If found, this mediated offer is returned to all parties to be considered against their current respective reservation utilities. If all parties accept the offer, then agreement is reached, otherwise the proposal and mediation procedures repeat, terminating at either eventual agreement or an imposed deadline. The proposal and mediation procedures are described in detail in the following two subsections.

Proposal Procedure

Lai and Sycara (2009) describe the linear proposing mechanism for two negotiating parties. According to this heuristic, when it is a party's turn to issue a proposal, it chooses to make an offer along the line that connects the two parties' previous best offers and has a utility equal to (or greater than) the proposer's current reservation utility. Although a party may not explicitly know its own utility function, it is reasonable to assume that it can identify a point with utility equal to its reservation utility, given that the utilities of points on the connecting line increase or decrease monotonically (Lai and Sycara 2009). This heuristic is an effective one because by approximating the point on the proposer's current reservation utility curve (or surface) that is the shortest distance from the opponent's previous best offer, the proposer is likely to be proposing to the opponent an offer of highest utility (Lai and Sycara 2009). In fact, Zheng et al. (Zheng et al.) have proven that such a shortest-distance proposal mechanism will eventually converge to an agreement, as long as the zone of agreement, *Z*, is non-empty.

We extend the linear proposing mechanism of Lai and Sycara (2009) to more than two parties. Specifically, we assume that the proposer makes an offer at a point along the line (in the multi-dimensional negotiation space) connecting the proposer's previous best offer and the *middle point* of the previous best offers of the other agents. Again, this point lies at the intersection of this line and the indifference surface representing the proposer's current reservation utility. We refer to this procedure as the *joint linear proposing* heuristic (Figure 1).

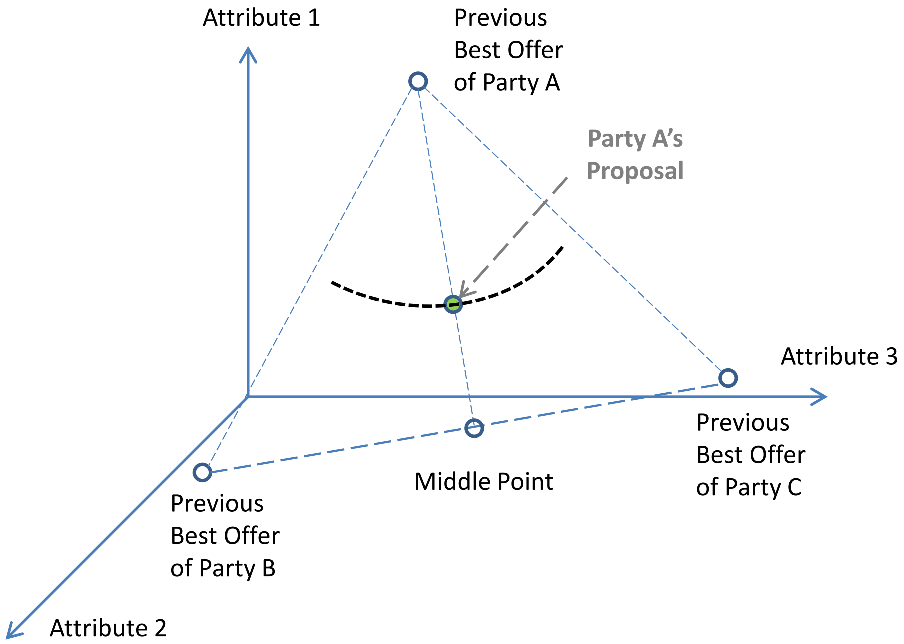

Figure 1: Schematic of joint linear proposing heuristic employed by Party A in negotiations with Parties B and C over attributes 1, 2, and 3. Black dashed curve represents the indifference (or iso-utility) curve at the level of Party A's current reservation utility.

We recognize that the linear proposing mechanism we employ may not fully representative of the actual process used to generate proposals in real-world negotiations. In fact, we are cognizant that in many contexts negotiators may be issuing proposals that do not conform to any utilitarian thought process. Nevertheless, this mechanism seems consistent with the heuristic of "splitting the difference," without allowing proposals that are below the proposer's current reservation utility. (It does assume, however, that the proposer is willing to concede all the way down to their reservation utility in each round, a point we come back to in the Discussion section.) Even if proposals are not actually being generated according to the joint linear proposing heuristic, such hypothetical offers provide a useful baseline for mediators to use as a starting point for subsequent queries. In fact, this is the context in which the linear proposing mechanism was described by Lai and Sycara (2009). Again, the heuristic has the significant practical advantage of not requiring parties to have explicit knowledge of their own utility function, nor any knowledge whatsoever of their opponents' utility functions.

Mediation Procedure

In our framework, the mediator has the task of finding a point that is mutually better for the responding parties than the offer just proposed. Lai and Sycara (2009) describe a procedure for two parties in a two-dimensional attribute space. We briefly review their procedure, highlighting extension to a situation involving >2 parties. For details and illustrative figures, we refer the reader to the original publication.

The procedure of Lai and Sycara (2009) relies on the concept of *ε-satisfaction*. A point $\boldsymbol{x}$ in the multi-attribute negotiation space is ε-satisfying for two parties if either: (1) there does not exist any point that is mutually better than $\boldsymbol{x}$ for both parties; or (2) all points that are mutually better than $\boldsymbol{x}$ are located within a distance ε of $\boldsymbol{x}$ in any attribute dimension. As $\varepsilon \rightarrow 0$, a point that is ε-satisfying in every dimension is Pareto optimal. In two attribute dimensions, to find an ε-satisfying point, the mediator begins by searching 'up' relative to the current offer. This means adding a value of ε to one attribute of the current proposal and issuing a series of queries to each party to find the corresponding value of the second attribute such that the party is indifferent between this new point and the existing offer. Depending on the relative values of the responses received from the parties, the mediator can determine whether a mutually better offer exists in the 'up' direction (see (Lai and Sycara 2009) for details). If so, the mediator replaces the current offer with a point that has ε added to the first attribute and averages the responses of the two parties regarding the second attribute and then searches 'up' again. If not, the mediator searches in the 'down' direction using a similar procedure. Once further mutual improvements cannot be found in either the ‘up’ or ‘down’ direction, the mediator declares the existing point ε-satisfying and offers it as the current mediated proposal.

To apply the mediation procedure to a multi-dimensional negotiation space, Lai and Sycara (2009) suggest projecting the multi-attribute proposal onto a two-dimensional surface (holding the value of the other attribute(s) fixed), executing the above mediation procedure, and then repeating in each of the other possible two-dimensional spaces. Figure 2 demonstrates this procedure graphically for three parties and three attributes. Point d represents Party A’s current proposal obtained by the joint linear proposing protocol described above. The previous best offer of each of the parties is first projected onto a surface represented by attributes 2 and 3 with the value of attribute 1 fixed at the corresponding value for point d (upper panel: points a, b, and c). In the resulting two dimensional space, the mediator then makes queries to parties B and C seeking Pareto improvements on the existing offer represented by point d (lower panel). The mediator first makes an ‘up’ search along the axis of attribute 3 by adding increment ε. By querying parties b and c, the mediator then identifies the values of attribute 2 that make the parties indifferent to point d (points e and f, respectively). Finally, because of the relative location of e and f, the intervening range is determined to be ε-satisfying, and the midpoint is offered as the mediated proposal.

As pointed out by Lai and Sycara (2009), the mediation procedure largely mimics the conventional process of utility function elicitation. However, by only querying and processing a few local points as the negotiation proceeds, rather than the entire negotiation space, the procedure significantly reduces the amount of mental and computational effort required.

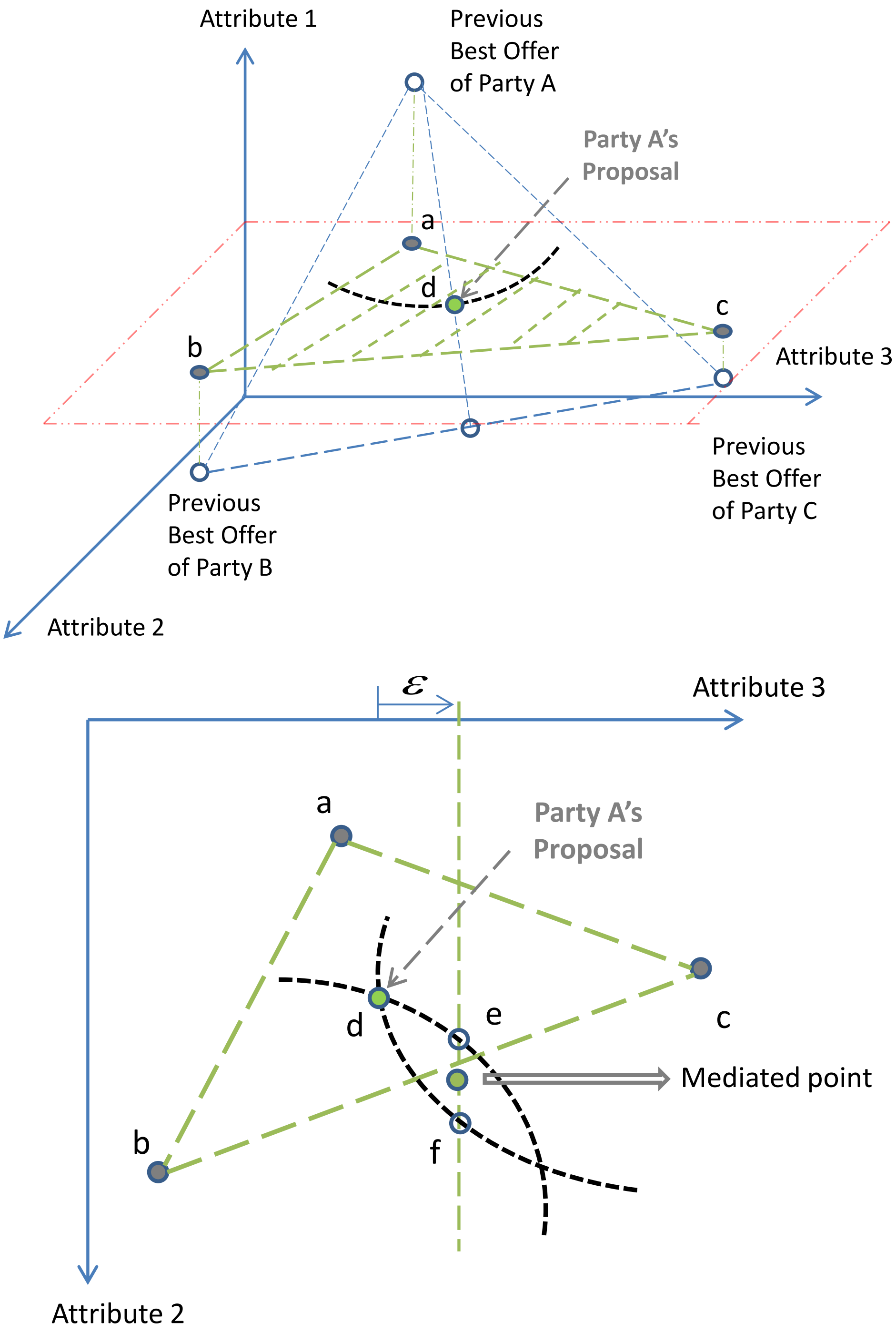

Figure 2: Schematic of mediation algorithm showing a three-dimensional proposal being projected onto a two-dimensional surface (upper panel), followed by a two-dimensional mediation procedure (lower panel). Black dashed curves in each panel represent indifference (or iso-utility) curves.

APPLICATION TO CLIMATE CHANGE NEGOTIATION

The urgency of the climate change problem suggests that, by hastening agreement and identifying treaties with near-Pareto optimal outcomes, an automated mediation procedure such as that described in the previous section could add value to international climate negotiations. In this section, we exemplify application of the procedure, adopting some strong simplifying assumptions regarding climate change forecasting, economic evaluation, and the negotiation process itself. Our goal is not to produce a fully realistic model, but rather to show in stylized form how automated mediation might work in practice and what results may be possible.

Negotiating Parties

For simplicity, we assume negotiations occur exclusively among the three parties who currently emit the majority of the world's greenhouse gases: China, the European Union (EU) and the United States (US). These parties negotiate over how much they should each reduce emissions relative to a baseline scenario (described below). All other nations are grouped into nine regions following Nordhaus's RICE model (Nordhaus 2010) and are assumed to either follow the emissions reduction commitment of one of the three negotiating parties or be exempt from emissions reductions (Table 1). These groupings generally follow the Umbrella and G77 associations (Kolstad and Toman 2001, Bang et al. 2009).

Region	**Emissions Control Group**
China Eurasia India Middle East Latin America	CHINA (G77)
US Japan Russia Other High Income	US (Umbrella)
EU	EU
Africa Other Asian	Exempt

Table 1: Assignment of twelve world nations/regions to emissions control groups.

Negotiation Attributes

In our model, the specific attributes being negotiated are how much each party *i* will reduce carbon dioxide emissions in 50 years relative to projected baseline emissions, $R_{i,50}$. The baseline emissions trajectory, $E^{b}_{i,t}$, represents a 'policy-free' future accounting for population and economic growth and a reduction in emissions intensity resulting primarily from scarcity effects on the price of fossil fuels. Our baseline trajectory for each region follows the specifications of RICE (Nordhaus 2010).

Negotiated reduction targets relative to the baseline are assumed to follow a time course represented by a cumulative Weibull curve (Meade and Islam 2006):

$$R_{i,t} = 1 - e^{(-\frac{t-t_0}{\lambda})^{K}} \quad (1)$$

where $R_{i,t}$ represents the reduction target relative to baseline for region *i* in year *t*, and t_0, λ and K represent shift, scale, and shape parameters, respectively. These parameter values are chosen to yield a negligible reduction in the first year ($R_{i,1} = 0.005$), the negotiated reduction target for 50 years in the future ($R_{i,50}$), and a near-complete curtailing of emissions at a distant point in time (Figure 3). Future emissions for each region $E_{i,t}$ are therefore calculated as:

$$= E_{i,t}\left(1 - R_{i,t}\right) \cdot E^{b}_{i,t} \quad (2)$$

and global emissions are calculated as the sum of the emissions of the 12 world regions.

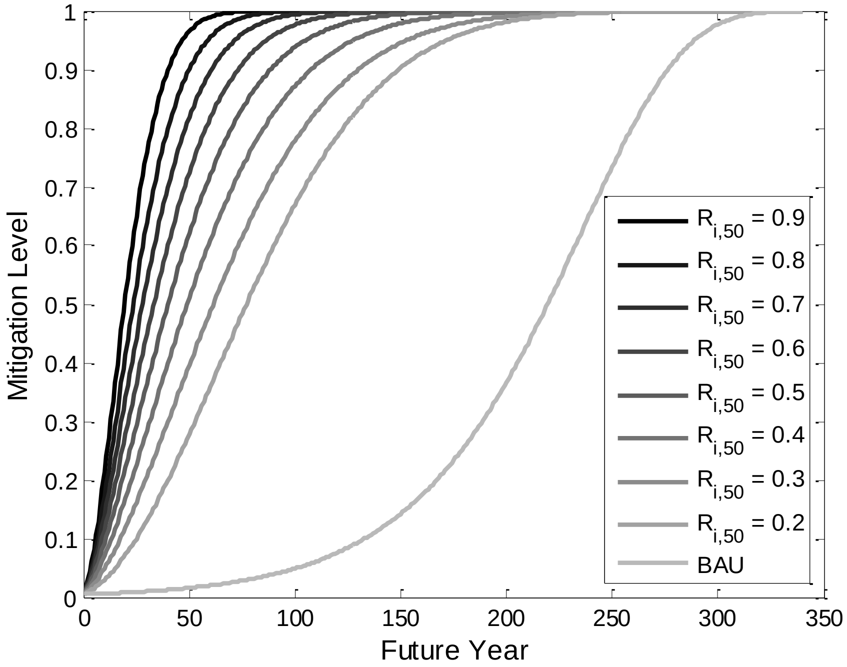

Figure 3: Emission reduction curves for representative values of $R_{i,50}$. BAU represents minimal 'business as usual' emission reductions induced by cost improvements in low-carbon technology not resulting from an international climate treaty (Nordhaus 2010).

Preference Structure

We assume that the utility of any given emissions reduction treaty to a negotiating party depends on its perceived net impact on future economic welfare. We therefore adopt a framework in which utility is represented by the ratio of the discounted gross domestic product (GDP) projected to result under the treaty as compared to discounted GDP without the treaty. With $\Lambda_{i,t}$ and $\Omega_{i,t}$ representing abatement and damage costs, respectively, (both expressed as a proportion of GDP) this gives:

$$U_{i,1}(R_{j=1,2,3;\tau}) = \frac{\sum_{t=\tau+1}^{\tau+100} (1+r)^{2010-t} \times \left[Q_{i,t}^{b} \times (1-\Lambda_{i,t} - \Omega_{i,t}) \right]}{\sum_{t=\tau+1}^{\tau+100} (1+r)^{2010-t} \times Q_{i,t}^{b}} \qquad (3)$$

where τ is the negotiation year, r is the discount rate, and $Q_{i,t}^{b}$ is business-as-usual GDP. Impacts are truncated at 100 years into the future, as per the policy termination timeframe of the Kyoto Protocol (Costa and Wilson 2000). For convenience, utility values are linearly transformed to span the [0, 1] range from the worst to the best possible outcomes for each party.

Economic and Climate Projections

All economic and climate change projections and assumptions come from RICE (Nordhaus 2010). The main relations are reproduced below, with details and parameter values given in an Appendix.

Emissions abatement costs for region *i* at time *t* are projected as:

$$\Lambda_{i,t} = B_{i,t} \times R_{i,t}^{2.8} \tag{4}$$

where $B_{i,t}$ is a coefficient reflecting the sensitivity of mitigation costs to emission reduction level and is assumed to be higher for developing regions and to decline with time.

Damage costs are estimated as:

$$\Omega_{i,t} = \omega_{L,i} \times T(t) + \omega_{Q,i} \times T(t)^2 \tag{5}$$

where T is global average temperature change above the pre-industrial value and $\omega_{L,i}$ and $\omega_{Q,i}$ are damage cost coefficients specific to each region.

In making temperature projections, negotiating parties are assumed to employ a reduced-form model of the carbon cycle. This model consists of two mass balance equations representing the flows of carbon between the atmosphere (M_{AT}) and biosphere and upper oceans (M_{UO}):

$$\begin{aligned} M_{AT}[t] &= E[t-1] + \varphi_1 M_{AT}[t-1] + \varphi_2 M_{UO}[t-1] \\ M_{UO}[t] &= (1-\varphi_1) M_{AT}[t-1] + (1-\varphi_2) M_{UO}[t-1] \end{aligned} \tag{6}$$

The climate system is then represented by two energy balance equations for the atmosphere and deep ocean that yield estimates of the change in temperature relative to pre-industrial levels:

$$T_{AT}[t] = T_{AT}[t-1] + \xi_1 \left(F_{2x} \cdot \log\left(\frac{M_{AT}[t]}{M_{AT}[t_{pre}]} \right) (\log 2)^{-1} - \left(\frac{F_{2x}}{S} \right) T_{AT}[t-1] - \xi_2 \left(T_{AT}[t-1] - T_{DO}[t-1] \right) \right) \tag{7}$$

$$T_{DO}[t] = T_{DO}[t-1] + \xi_3 \left(T_{AT}[t-1] - T_{DO}[t-1] \right) \tag{8}$$

where ξ_1, ξ_2, and ξ_3 are parameters dictating the rates of temperature adjustment in the atmosphere and of heat transfer, and parameters F_{2X} and S characterize the effect of a doubling of atmospheric CO_2 concentration on radiative forcing and global temperature, respectively.

Reservation Utility

We assume that the reservation utility for each party is time-dependent and is represented by its utility for a scenario in which all regions follow the BAU ('no agreement') emissions trajectory going forward. No party should rationally accept an agreement that is worse than its current BATNA utility. Because the climate damages for each region have differing sensitivity to global temperature change, the reservation utilities of the negotiating parties differ. A lower or more rapidly declining reservation utility indicates a greater willingness to reach agreement.

Negotiation Procedure

We assume that the three negotiating parties meet annually and, at each meeting, the EU, China, and the US sequentially have the opportunity to make proposals. Proposals are made according to the joint linear proposing heuristic. After each proposal, the mediator uses the query process described to try to improve the current proposal for the two responders. In responding to the mediator's queries, each party is assumed to make projections of the climate and economic implications of the current offer using eqs. (4) through (8) and then convert these projections into relative preferences using eq. (3). We use a value of ε=0.01 as the basis for the mediator's queries and limit the number of queries in each mediation to three in order to cap the level of mediator and responder effort required in each round. In sensitivity analysis, we evaluate the effect of changing this limit.

Upon receiving a mediated proposal, each party again makes the necessary climate and economic projections and compares the resulting utility of the proposal to its current reservation utility. Any proposal with greater utility than the reservation value will be accepted. If all parties accept the offer, then agreement is reached and negotiation ends. Otherwise the proposal and mediation procedures repeat until each party has made a proposal in that year. If no agreement is reached, then negotiations are suspended until the following year. Proposals made in any given year are remembered and used as the basis for proposals in the subsequent year. Emissions and climate change proceed while negotiations occur, changing both the reservation utility and the utility of any agreement moving forward. Economic and climate conditions at the start of negotiation correspond to those in the year 2010. For illustration purposes, we assume that each party starts the negotiations with the position that its opponents should cut emissions dramatically ($R_{i,50}$ = 0.8), while it must do very little itself ($R_{i,50}$ = 0.05). We then monitor the subsequent negotiation proceedings, both with and without the use of automated mediation and compare the resulting agreements to the Nash bargaining solution. The *Nash bargaining solution* at time t is the negotiation point $\boldsymbol{x}$ that maximizes the product across all parties of the difference between their utility for $\boldsymbol{x}$ and their current reservation utility:

$$\prod_{i=1}^{N} \left(U_{i,t}(x) - ru_i(t) \right) \tag{9}$$

This expression is referred to as the *Nash product*, and the corresponding Nash bargaining solution ensures that each party gets its reservation utility (i.e., the utility associated with no agreement) in addition to an equal share of the benefits yielded by the agreement (Nash 1950). In our setup, parties are not fully cognizant of their own utility functions and do not share information on their preferences. Therefore, they cannot find the Nash bargaining solution. Nevertheless, it provides a useful reference point as a ‘fair’ Pareto optimal agreement.

All calculations were implemented in MATLAB (R2013a) and code is available from the authors upon request.

Results

With no mediation, the negotiating parties reach agreement in the ninth year of negotiations, settling on 50-year reduction targets of R_{50} = (0.50, 0.60, 0.55) for China, the EU, and the US, respectively (Figure 4, left). Over the time course of negotiations, each proposer's own proposed emission reduction targets increases over time, while its opponents' proposed targets decrease over time. This is a result of the compromising joint linear proposal heuristic, as well as a reservation utility that decreases over time (Figure 4, right). When formulating a proposal, the proposing party makes an offer that follows its own reservation utility curve while meeting its opponents part way. In this way, the proposal and reservation utility curves slowly approach one another until they converge for all parties and agreement is reached. This agreement, which assumes that all regions, including those listed in Table 1, adhere to their committed targets, has corresponding utility values to China, the EU, and the US, respectively, of $U_{i,9}$ = (0.909, 0.869, 0.891). This is the net result of a projected temperature change of 2.6ºC by 2100, resulting in damage costs at that point of 1.08, 1.10, and 0.98% of GDP for China, the EU, and the US, respectively. Expenditures on emissions abatement by the three parties at that point are projected to reach 0.83, 0.92, and 0.76% of GDP, respectively.

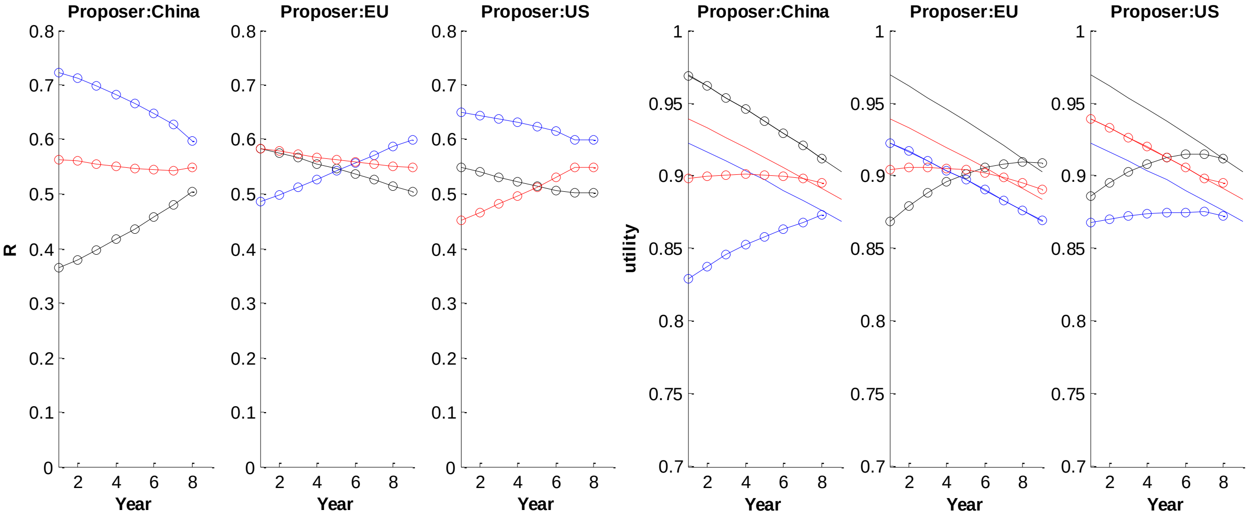

Figure 4: Left: Proposed emission reduction targets in each round of negotiations without any mediation. Right: Corresponding utilities of proposals for each negotiating party (pointed lines) compared to reservation utilities (solid lines). Parties agree to the proposal made by the EU in year 9. Black=China; Blue=EU; Red=US.

The Nash bargaining solution at year 9 would involve a reduction target of R_{50} = (0.33, 0.16, 0.28) with $U_{i,9}$ = (0.914, 0.889, 0.896). The fact that the parties reach agreement at an emission reduction target that is distant from the Nash bargaining solution indicates that, under our assumptions, a non-mediated negotiation process may converge to a treaty that satisfices for each party, but has plenty of room for improvement.

Employing our mediation protocol, agreement is reached in year 4 with 50-year reduction targets for China, the EU, and the US of R_{50} = (0.34, 0.20, 0.26), respectively. Under mediation, the progression of proposals is less predictable (Figure 5, left). The mediated offers that evolve from a party's proposal in each round may increase, decrease, or stay the same over time, depending on what would generate a Pareto improvement. The mediated offers are also seen to span a much larger range, thereby exploring the negotiation space more broadly. After mediation, offers also track each proposer's reservation utility curve less closely (Figure 5, right) and can remain well above the curve, so long as opponents are making proposals that deliver a utility greater than the BATNA.

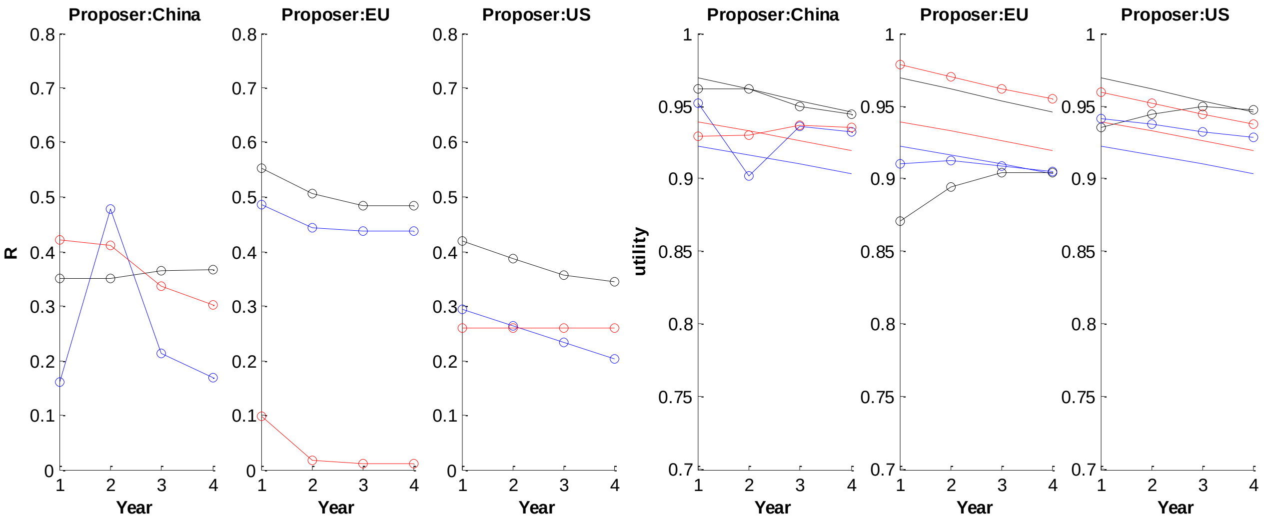

Figure 5: Left: Mediated emission reduction target proposals in each round of negotiations. Right: Corresponding utilities of mediated proposals for each negotiating party (pointed lines) compared to reservation utilities (solid lines). Parties agree to the mediated proposal of the US in year 4. Black=China; Blue=EU; Red=US.

The agreement after mediation in year 4 (Figure 6) has corresponding utility values to China, the EU, and the US, respectively, of $U_{i,4}$ = (0.946, 0.927, 0.939). Associated abatement costs in 2100 represent 0.52, 0.26, and 0.31% of GDP for China, the EU, and the US, respectively, while damage costs resulting from a projected temperature change in 2100 of 2.96ºC are projected to comprise 1.59, 1.69, and 1.50% of GDP. These values can be compared to projected BAU damage costs in 2100 of 2.12, 2.30, and 2.05% of GDP (Figure 7).

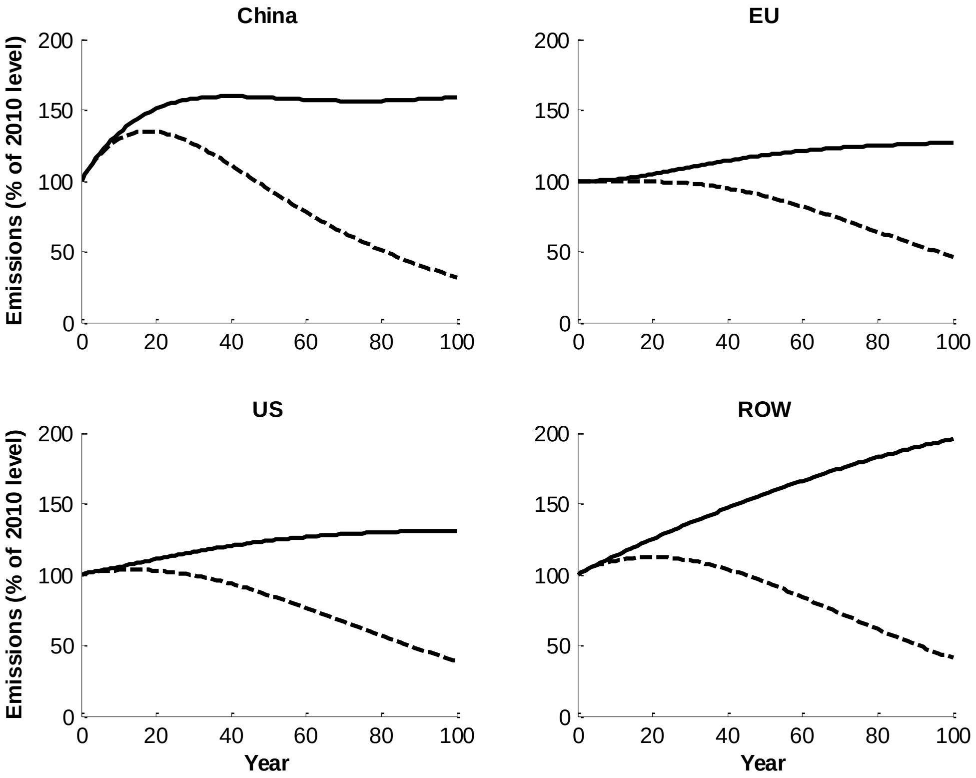

Figure 6: A comparison of emissions trajectories under the BAU 'no-agreement' scenario (solid black) against the targets of the mediated agreement (dashed black). ROW = 'Rest of World' and represents the total emissions of all non-negotiating parties.

The agreement reached under mediation is much closer to the Nash bargaining solution in year 4 of R_{50} = (0.28, 0.13, 0.23) with corresponding utilities of $U_{i,9}$ = (0.950, 0.920, 0.928). Thus, not only does the mediation procedure lead to earlier agreement, but also to a treaty that is closer to Pareto optimality.

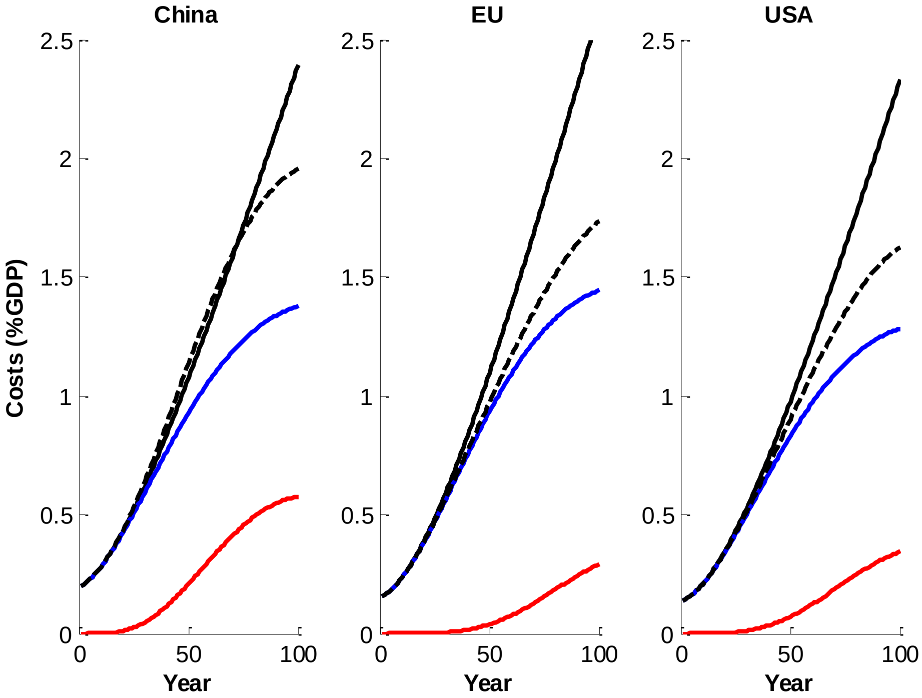

Figure 7: A comparison of damage costs under the BAU 'no-agreement' scenario (solid black) against total costs associated with the mediated agreement (black dashed). Costs associated with the mediated agreement are further subdivided into damage costs (blue) and abatement costs (red).

As a next step, we investigate the sensitivity of model results to the maximum number of mediator queries allowed to be posed to each party in each mediation step. Results (Figure 8) show that even with only one query to each party, the mediation algorithm can reduce the time to agreement by 33% and substantially increase the utility of each party for the agreement. A solution with utilities to each party that are very close to the Nash bargaining solution at four years can be reached with only four queries to each party in each mediation step. This confirms that the effort level required of negotiators by the automated mediation procedure can be minimal while yielding substantial gains.

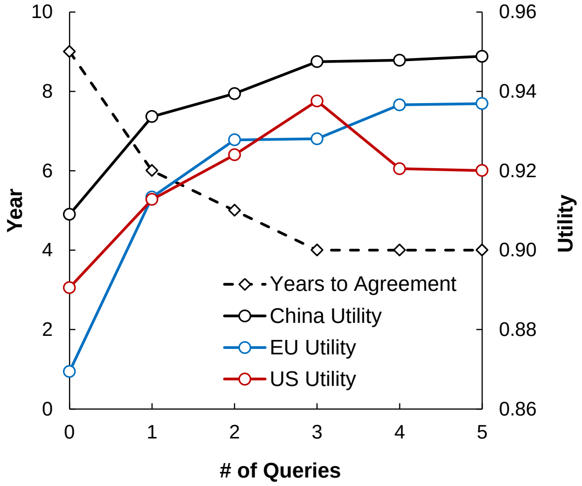

Figure 8: Sensitivity of the timing (dashed line, left axis) and utility (solid lines, right axis) of agreement to the maximum number of mediator queries allowed to be posed to each party in each mediation step.

DISCUSSION

An automated mediation procedure, such as the type introduced by Lai and Sycara (2009) and modified here, can help participating parties to reach agreement sooner and with outcomes that are closer to Pareto optimality. Thus the parties can take comfort in knowing that opportunities for further mutual improvement are not being overlooked. Our climate treaty example also shows that the additional effort level required, in the form of answers to mediator queries, need not be substantial, with 1 to 3 queries per mediated proposal being sufficient to yield large improvement.

Our model employs assumptions regarding climate change forecasting and economic evaluation that are consistent with the state-of-the-art in integrated assessment modeling (Nordhaus 2010). Yet, we do not want to claim that our specific predictions regarding the timing and nature of agreement (either with or without automated mediation) are accurate. For example, we assume a simple and straightforward joint linear proposing heuristic for generating each party's proposals in each round. This step is not part of our mediation procedure, but rather is intended to serve as a stand-in for actual proposals made by human negotiators employing whatever principles they choose. Negotiators that are less willing to 'split the difference' with opponents will likely be involved in lengthier negotiations (both with and without mediation) than those suggested by our model. Similarly, we assume that responders accept any proposal with a utility greater than its reservation utility, or BATNA. Further, we take this BATNA to correspond to the BAU emissions trajectory. In reality, a negotiating party may demand that any acceptable treaty yield benefits that are more than just marginally better than the status quo. This could be handled by adding a specified increment to a party's BATNA utility, although such arbitrary modifications would only reduce the transparency of our example.

In our Introduction, we reasoned that basing international climate policy primarily on the identification of a 'fair' burden-sharing scheme is unlikely to be successful because each party will judge 'fairness' largely according to whether its own interests are being served (Lange et al. 2010). Nevertheless, it may make sense to add a component to our utility function to represent each party's aversion to inequity in addition to the GDP-based component representing pure self-interest. Fehr and Schmidt (1999) suggest various forms for such a component which represent the disutility a party feels when an opponent ends up relatively better (or worse) off than itself as a result of the negotiated agreement. Such a component could create some interesting negotiation dynamics, and we leave exploration of this idea for future work.

In other publications, we have argued that considerations of uncertainty and risk attitude should be reflected in utility-based climate policy assessments. In particular, the large uncertainty in climate damage forecasts (Gerst et al. 2010), combined with strong public aversion to economic risk (Ding et al. 2012), leads to a high risk premium associated with greenhouse gas emissions mitigation (Gerst et al. 2013). For simplicity, we ignore this effect in the present model, but its absence is likely to lead to an underestimate of the emissions reduction targets that parties are willing to agree to.

Other simplifying assumptions have to do with the negotiation process itself. In reality, the process is not so clearly dominated by the EU, China, and the US, and the procedure is not as clearly structured as the annual propose-and-response protocol we assume. The attributes being negotiated are also certainly more complex than 50-year emissions reduction targets, each implying a smooth and continuous S-shaped curve of reductions over the next 200+ years. We also assume that all nations within each emissions control group follow the commitment of its lead negotiating party. By ignoring the defections that are implied by the 'tragedy of the commons' concept, this assumption likely leads to over-optimism about the chance and magnitude of an emissions reduction agreement (ref)

Despite – or perhaps because of – the real-world complexities of international climate negotiation, an automated mediation process can add value by encouraging parties to be clear about their own assumptions and preferences. While many nations are already using mathematical models to support their negotiation practices (Sterman et al. 2013), most models focus on forecasts of climate change and associated damages. We are not aware of any that provide support in how to make, or respond to, actual proposals in a way that facilitates eventual agreement. To use our mediation procedure, parties only need to be able to perform basic preference comparisons, stating whether one multiattribute offer is more, less, or equally preferred to another. Fully-articulated utility functions are not required. Yet, even this degree of preference consideration is likely to stretch negotiators to think explicitly about the tradeoffs between abatement cost and climate damages and about the relative impacts across nations.

References

Bang, G., G. Heggelund, and J. Vevatne. 2009. Shifting Strategies in the Global Climate Negotiations. Fridtjof Nansen Institute (FNI).

Binmore, K., A. Rubinstein, and A. Wolinsky. 1986. The Nash bargaining solution in economic modelling. The RAND Journal of Economics:176-188.

Blok, K., N. Hohne, A. Torvanger, and R. Janzic. 2005. Towards a Post-2012 Climate Change Regime.

Chatterjee, K., and W. Samuelson. 1983. Bargaining under incomplete information. Operations Research 31:835-851.

Costa, P. M., and C. Wilson. 2000. An equivalence factor between CO2 avoidedemissions and sequestration-description andapplications in forestry. Mitigation and Adaptation Strategies for Global Change 5:51-60.

Ding, P., M. D. Gerst, A. Bernstein, R. B. Howarth, and M. E. Borsuk. 2012. Rare Disasters and Risk Attitudes: International Differences and Implications for Integrated Assessment Modeling. Risk Analysis.

Endriss, U. 2006. Monotonic concession protocols for multilateral negotiation. Pages 392-399 *in* Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems. ACM.

Fatima, S. S., M. Wooldridge, and N. R. Jennings. 2006. Multi-issue negotiation with deadlines. J. Artif. Intell. Res.(JAIR) 27:381-417.

Fehr, E., and K. M. Schmidt. 1999. A theory of fairness, competition, and cooperation. The Quarterly Journal of Economics 114:817-868.

Gerst, M. D., R. B. Howarth, and M. E. Borsuk. 2010. Accounting for the risk of extreme outcomes in an integrated assessment of climate change. Energy Policy 38:4540-4548.

Gerst, M. D., R. B. Howarth, and M. E. Borsuk. 2013. The interplay between risk attitudes and low probability, high cost outcomes in climate policy analysis. Environmental Modelling & Software 41:176-184.

Gupta, J. 2010. A history of international climate change policy. Wiley Interdisciplinary Reviews: Climate Change 1:636-653.

Harsanyi, J. C., and R. Selten. 1972. A generalized Nash solution for two-person bargaining games with incomplete information. Management Science 18:80-106.

Jennings, N. R., P. Faratin, A. R. Lomuscio, S. Parsons, M. J. Wooldridge, and C. Sierra. 2001. Automated negotiation: prospects, methods and challenges. Group Decision and Negotiation 10:199-215.

Kolstad, C. D., and M. Toman. 2001. The economics of climate policy. Handbook of Environmental Economics 2.

Kraus, S. 2001. Automated negotiation and decision making in multiagent environments. Lecture notes in computer science:150-172.

Lai, G., and K. Sycara. 2009. A generic framework for automated multi-attribute negotiation. Group Decision and Negotiation 18:169-187.

Lange, A., A. Loschel, C. Vogt, and A. Ziegler. 2010. On the self-interested use of equity in international climate negotiations. European Economic Review 54:359-375.

Lin, R., S. Kraus, J. Wilkenfeld, and J. Barry. 2008. Negotiating with bounded rational agents in environments with incomplete information using an automated agent. artificial Intelligence 172:823-851.

Meade, N., and T. Islam. 2006. Modelling and forecasting the diffusion of innovation-A 25-year review. International Journal of Forecasting 22:519-545.

Nash, J. 1951. Non-cooperative games. The Annals of Mathematics 54:286-295.

Nash, J. F., Jr. 1950. The bargaining problem. Econometrica 18:155-162.

Nordhaus, W. D. 2010. Economic aspects of global warming in a post-Copenhagen environment. Proceedings of the National Academy of Sciences 107:11721-11726.

Raiffa, H. 1982. The art science of negotiation. Harvard University Press.

Ringius, L., A. r. Torvanger, and A. Underdal. 2002. Burden sharing and fairness principles in international climate policy. International Environmental Agreements 2:1-22.

Rubinstein, A. 1982. Perfect equilibrium in a bargaining model. Econometrica: Journal of the Econometric Society:97-109.

Sjo'stedt, G., and A. M. Penetrante. Strategic Facilitation of Climate Change Negotiations. Climate Change Negotiations: A Guide to Resolving Disputes and Facilitating Multilateral Cooperation:3.

Stenlo, L. 1972. Mediation in International Negotiation. MalmÃ¶: Nordens Boktryckeri.

Sterman, J. D., T. Fiddaman, T. Franck, A. Jones, S. McCauley, P. Rice, E. Sawin, and L. Siegel. 2013. Management flight simulators to support climate negotiations. Environmental Modelling & Software 44:122-135.

Zheng, R., N. Chakraborty, T. Dai, and K. Sycara. 2013. Automated Multiagent Negotiation on Multiple Issues with Private Information.