

The Effects of Goal Revelation on Computer-Mediated Negotiation

Ya'akov Gal
MIT CSAIL
Harvard Univ. SEAS

Sohan D'souza
British Univ. in Dubai

Philippe Pasquier
Simon Fraser Univ.

Iyad Rahwan
British Univ. in Dubai
Univ. of Edinburgh

Sherief Abdallah
British Univ. in Dubai
Univ. of Edinburgh

Abstract

This paper studies a novel negotiation protocol in settings in which players need to exchange resources in order to achieve their own objective, but are uncertain about the objectives of other participants. The protocol allows participants to request each other to disclose their interests at given points in the negotiation. Revealing information about participants' needs may facilitate agreement, but it also exposes their negotiation strategy to the exploitation of others. Empirical studies were conducted using computer-mediated negotiation scenarios that provided an analogue to the way goals and resources interact in the world. The scenarios varied in the individual positions and interests of participants, as well as the dependency relationships that hold between participants. Results show that those who choose to reveal their underlying goals outperform negotiators in the same setting that use a protocol that forbids revelation. In addition, goal revelation has a positive effect on the aggregate performance of negotiators, and on the likelihood to reach agreement. Further analysis show goal revelation to be a cooperation mechanism by which negotiators are able to identify acceptable agreements in scenarios characterized by few socially (Pareto) beneficial outcomes.

Introduction

Goals and incentives are key determinants of human behavior, but in many negotiation scenarios there is lack of information about the underlying interests of participants. Often, this prevents the parties from reaching a beneficial agreement. Consider a bank who is offering to purchase the majority of shares of a struggling company in return for potential job cuts. The unions may not allow the company to accept the offer because they refuse to agree to layoffs. However, if the bank discloses that it is committed to keeping the company afloat, the unions may agree to the buy-out if layoffs are minimal. On the other hand, revealing goals is often associated with a cost. Having realized that the bank does not intend to liquidate the company, the unions may demand no job cuts.

This work studies the trade-offs associated with different negotiation protocols in settings where self-interested parties lack information about each other's aims. We consider strategic settings which require an agreement on the allocation of scarce resources among self-interested parties. Participants take turns proposing take-it-or-leave-it deals to each other under time constraints, and communication is associated with a cost. With no certain knowledge of each other's goals, the offers of participants serve as a "noisy signal" to their true objectives. It is difficult to locate efficient trades for both parties in such conditions, either because participants may request more than they need, or because there are simply too many combinations of possible agreements to try out under time constraints.

In these conditions, revealing the objectives of one or more of the participants may facilitate agreement, because the addi-

tional information narrows the "search space" of possible offers, and may reveal new avenues of negotiation that were not known before. However, it is not obvious that the revelation of information by either party will necessarily improve the result of the negotiation. Goal revelation is potentially costly, because it exposes the revealing party's position and negotiation strategies. For example, a participant that revealed its goals could have exposed itself to be very selfish in its past offers, and this may adversely affect the quality of the deals they are offered in the future.

This paper proposes a novel interest-based negotiation protocol in task settings, in which parties can prompt each other to reveal their goals at fixed points within the negotiation process. This protocol is inspired by recent negotiation protocols that allow participants to exchange information about beliefs, goals or social aspects (Rahwan et al., 2003). We compare this interest-based protocol with an alternative position-based protocol where goal revelation is not allowed. We measured people's behavior under each of these protocols using a computer-mediated testbed comprising a conceptually simple game in which players negotiate and exchange resources to enable them to achieve their individual goals. This testbed has been used previously to analyze the decision-making strategies people deploy when interacting with computers, and the comparison of these strategies with those that people deploy when interacting with other people (Gal, Pfeffer, Marzo, & Grosz, 2004). The advantage of this testbed for studying interest-based negotiation is two-fold: First, it presents a realistic analogue to the ways in which goals, tasks and resources interact in real-world settings, but it abstracts away the complexities of specific domains. Second, it supports transparent, anonymous communication between subjects who are interacting together in laboratory conditions, avoiding experimenter effects and face-to-face communication.

We conducted experiments in which different subjects interacted with each other using either interest- or position-based protocols on the same set of negotiation scenarios. These scenarios varied in the dependency relationships that hold between players (i.e., who needs whom), as well as the number of integrative (mutually beneficial) exchanges. Results show that goal revelation using interest-based negotiation protocols leads to a higher likelihood of agreement, and a significant increase in benefit to the revealing player, as compared to the benefit obtained using the position-based protocol. In addition, using the interest-based protocol itself has a positive effect on the social benefit to both parties which is significantly higher than the social benefit obtained using the position-based protocol. Further analysis revealed that interest-based negotiation is essentially a mechanism by

which dependent players who reveal their goals are assisted by independent players, while not incurring a loss to their own benefit, as compared to using position-based protocols.

There are few works offering an empirical analysis of people’s negotiation strategies in repeated interactions. Rubinstein (1985) has provided a theoretical model for prescribing negotiating strategies in such settings that are optimal under certain conditions (e.g., participants are rational and consistent in their beliefs about each other’s objectives). Work in the psychological literature about strategic interaction has focused on specific domains (e.g., seller-buyer disputes (G. Loewenstein, Bazerman, & Thompson, 1989), mid-east peace talks (Atran, Axelrod, & Davis, 2007)) or completely abstract settings such as the Prisoners Dilemma. Loewenstein and Brett (2007) conducted a study which studied how goal framing prior to the negotiation procedure affects strategy revision. None of these works have compared the effects of goal revelation directly within repeated negotiation. Work by Heiskanen, Ehtamo, and Hamalainen (2001) has found that people minimize the amount of private information they reveal in negotiation to avoid weakening their positions, while Vorauer and Claude (1998) observes that negotiators overestimate their partners’ beliefs about their own goals in certain conditions. We show that when using a mediated protocol for goal solicitation and revelation, negotiators are willing to disclose private information to others and that this allows them to reach mutual beneficial agreements.

Work in automated negotiation in Artificial Intelligence (AI) has proposed algorithms for argumentative strategies which support or attack the different positions of parties in a negotiation (Kraus, Sycara, & Evenchik, 1998). These algorithms have been used by computational agents and several works have studied conditions under which such strategies outperform position-based protocols (Kakas & Pavlos, 2006; Rahwan, Pasquier, Sonenberg, & Dignum, 2007). This work directly extends these studies by showing that argumentative-type protocols are advantageous to people.

The Colored Trails Game

Colored Trails (CT) is a testbed developed for investigating the decision-making that arises in task settings, where the key interactions are among goals, tasks required to accomplish those goals, and resources needed to perform the tasks¹. The empirical investigations described in this paper used a particular configuration of CT that is played by 2 players on a 5x5 board of colored squares. Each player had a designated goal square and a piece on the board, initially located in one of the non-goal squares. At the onset of a CT game, players are issued a set of 7 colored chips chosen from the same palette as the squares. To move a piece into an adjacent square a player must turn in a chip of the same color as the square. Chips represent resources in CT, and at the heart of the game is players’ ability to negotiate over these resources. Thus, chips may be exchanged by the players, and the conditions of

¹CT can be downloaded at <http://www.eecs.harvard.edu/ai/ct>



(a) Main and Pending Offer Panels



(b) Decision-support tool

Figure 1: Snapshot of a CT Game using an Interest-Based Negotiation Protocol

exchange may be varied to model different decision-making situations. The colors of the chips issued to players and the tiles on the board are drawn from a common 4-color palette. Players had full view of the board and each others’ chips and positions, but they could only see their own goal location.

A CT game comprises three phases. In the *communication phase*, players alternated between one of two roles: *proposer* players could offer some subset of their chips to be exchanged with some subset of the chips of responder players; *responder* players could in turn accept or reject proposers’ offers. If no offer was made, or if each offer was declined, then both players are left with their initial allocation of chips. The game controller prevents players for offering chips they do not have, or asking for chips the other does not have. In the *exchange phase*, chip exchanges were enforced by the game controller (if an agreement was reached). Finally, in the *movement phase*, the game controller automatically moved both players as close as possible to the goal square. The scoring function for players depended solely on their own performance: 100 points for reaching the goal; 10 points for each tile left in a player’s possession; 15 points deducted for any square in the shortest path between a player’s final position and goal-square (in case the goal was not reached). These parameters were chosen so that getting to the goal was by

far the most important component, but if a player could not get to the goal it was preferable to get as close to the goal as possible. The score that each player received if no offer was made was identical to the score each player received if the offer was rejected by the responder. We refer to this score as the *no negotiation alternative* and to the score that each player received if the offer was accepted by the responder as the *proposed outcome* score.

Snapshots of the CT graphical user interface (GUI) for the interest-based protocol of one of the games used in the experiment is shown in Figure 1. The Main “Window” panel, shown in Figure 1a, includes the board game, the goal square, represented by an icon displaying the symbol G_{me} , and two icons, “me” and “O”, representing the location of the two players on the board at the onset of the game.² The bottom part of the “Main Window” panel, titled “chips”, shows the chip distributions for the players. In the game shown here, the “me” player can get to the goal square, using the path that is outlined on the board, but the “O” player is lacking the chips to get to the goal (note that O’s goal is not shown here). The “me” player has received an offer asking it for 1 purple chip in return for 1 green chip. A proposer uses the “Propose Exchange” panel, to make an offer to a responder, or to ask for the other’s goal. The “Path Finder” panel, shown in Figure 1b, provides decision support tools to be used during the game. It displays a list of path suggestions to the goal, the missing chips required to fulfil a potential path, the surplus chips left over once a potential path has been fulfilled, and the best position the agent can reach relative to its scoring function. These paths are optimized for a given chip distribution and player, as queried by the player, such that they represent the best route given a player’s objectives. Agents can view this information for the chip set that is currently in their possession, or for any hypothetical chip set.³

Analogy with Task Settings

CT provides a realistic analog to task settings, highlighting the interaction among goals, tasks required to achieve these goals, and resources needed for completing tasks. Chips correspond to agent capabilities and skills required to fulfill tasks. Different squares on the board represent different types of tasks. A player’s possession of a chip of a certain color corresponds to having the skill available for use at a time. Not all players possess chips in all colors, much as different agents vary in their capabilities. Traversing a path through the board corresponds to performing a complex task whose constituents are the individual tasks represented by the colors of each square. There can be several paths on the board to get to the goal, as there are several ways of completing a task. It has been shown that people are more likely to engage in cooperative behavior when using this game than when using completely abstract representations such as payoff matrices (Gal,

Grosz, Shieber, Pfeffer, & Allain, 2007).

Interest- and Position-Based Protocols in CT

We now describe the implementation of interest- and position-based protocols in CT. In both protocols, neither player can see the goal of the other at the onset of the game, and players are randomly allocated as proposers or responders. In the communication phase, a proposer can make an offer to the responder, as shown in Figure 1a.

In the position-based protocol, once a responder receives an offer, it can accept it, in which case the offer is realized, both players automatically move towards the goal, and the game ends. If the responder rejects the offer, the game controller reverses the players’ roles, and the new proposer player (formerly the responder) can make an offer to the new responder player (formerly the proposer). A state-based representation of this protocol is shown in Figure 2a.

The interest-based protocol is an extension of the position-based protocol that allows players, in a controlled fashion, to ask about, and reveal, their goals. Once a responder receives an offer from the proposer, it has the option to ask the proposer for its goal, in addition to rejecting or accepting the offer. If the responder chooses not to ask for the goal, the game proceeds as in the position-based negotiation case. If the responder chooses to ask the proposer for its goal, the proposer now has the option to agree to reveal its goal, or to make another offer to the responder, which is effectively a rejection of the revelation request. Responders may ask proposers for their goals numerous time, but once a goal is revealed, it cannot be asked about, or revealed, again. A state-based representation of this protocol is presented in Figure 2b.

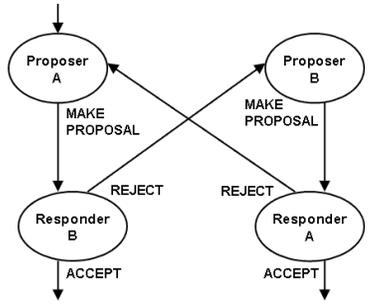
In both conditions, the players are allowed a maximum of 4 minutes to interact in the communication phase until they reach an agreement. If no offer is accepted, the communication phase terminates with no agreement. In both cases, the movement phase is initiated and both players are moved towards their goals given the result of the negotiation.

Empirical Methodology

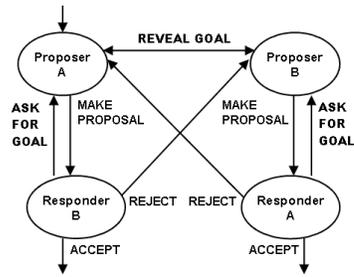
We refer to the session involving the interest-based negotiation protocol as the IBN condition, and the session involving position-based negotiation protocol as the PBN condition. Twenty-two subject participated in the experiment, drawn from a pool of students and adults residing in the Boston area. Twelve people participated in the IBN condition while ten people participated in the PBN condition. Each person was given an identical 30 minute tutorial on CT, and played 3 practice rounds. Each subject was identified by a serial number, was seated in front of a terminal for the entire length of the experiment, and could not see or speak to any of the other participants. A central server was responsible for matching up the participants at each round and for keeping the total score for each subject in all rounds of the experiment. Each subject participated in a series of CT games, but no subject was paired up with any other subject more than once, and subjects were not told about the identity of their counterparts.

²CT colors have been converted to grey scale in this figure.

³Players can query the path finder about the other player only if it has revealed its goals, and never query the path finder in position-based protocol.



(a) Position-Based Negotiation Protocol (Goal Revelation disallowed)



(b) Interest-Based Negotiation Protocol (Goal Revelations allowed)

Figure 2: Two Alternative Protocols for Repeated Negotiation

Participants were paid in a manner consistent with their aggregate scores in all of the games they played. Between each game, players engaged in a neutral activity which did not affect their payment (answering questions about their mood), designed to minimize the effects of past games on their future performance. In addition, subjects scores were not revealed at any point during the experiment.

The games in each condition were generated to meet the following criteria:

- It was not possible for *both* players to reach their respective goals independently. This ensures that it is worthwhile for players to negotiate.
- Either player can reach its goal with the other’s help, but it may not be possible for both players to reach the goal. This ensures that in any game, it is potentially possible for at least one player to reach its goal.
- The players and their respective goals must be in either the same row or column of the board. This constraint allows the player one shortest path to the goal, and makes it easier to consider alternative paths that may be realized due to goal revelation.

For each game, we recorded the board and chip settings, as well as the actions of both players and their associated scores in the game.

We hypothesized that players’ ability to reveal their goals in CT will lead them to identify alternative plans on the board,

different from the paths they were meaning to take before the goal revelation. Consequently, chips that were meant to be used for the original path may be “released” and potentially help the other player. We expected that this information will help players reach more beneficial agreements for both parties.

Results

	Goal Revelations 1, 2	All Games
Revealing Player	41 / 23	38 / 38
Soliciting Player	8 / 9	34 / 32
Combined	49 / 32	72 / 61

Table 1: Average benefit in IBN/PBN condition for different number of goal revelations (significant difference shown in bold)

We base the following results on the same 65 games that were played in both conditions. In 14 of these games players were co-dependent (both players needed each other to get to the goal) and in the other 51 games one of the players needed the other player. We refer to a participants that queried their counterpart about their goal as a “goal solicitor,” and to those players that subsequently revealed their goal as “goal revealers”. We will also use the term “goal revealers” in the PBN condition, despite the fact that there is no revelation in this condition. In the PBN condition, this term refers to those players who align with goal revealers in the equivalent set of games in the IBN condition (and similarly for goal solicitors).

Table 1 shows the average benefit to revealers and solicitors in games played in the IBN condition (left entry) and the PBN condition (right entry). The benefit to a player in a game is defined as the proposed outcome score minus the no negotiation alternative score. If no agreement is reached, a player’s benefit is zero. The results are measured with respect to the total set of games played in both conditions (column marked “All Games”) and with respect to the games in which one or two goals were revealed in the IBN condition (column marked “Goal Revelations”). Recall that these same games were also played in the PBN condition.

As shown in the table, the combined average benefit for players in the IBN condition was 49, which is 17 points greater in cases where there was at least one goal revelation ($SE = 5$, paired t-test $t(29) = 1.7, p = 0.04$). There was an insignificant 5 points point difference between conditions in the combined benefit for those games in which no revelations occurred ($SE = 2$, paired t-test $t(34) = 0.5, p = 0.28$). We can conclude that revelation itself has a positive effect on players’ social benefit. This is because there was a combined advantage to players in the IBN condition when the number of revelations was greater than zero, but not when this number equalled zero. In addition, it turns out that this advantage is significant enough to affect players’ total performance. As

	IBN	
	agreement	failure
PBN agreement	11	7
PBN failure	16	31

Table 2: Pairwise Agreement Ratio

	Goal Revelations		All Games
	0	1, 2	
indep. player	19 / 0	-9 / -15	15 / -2
dep. player	49 / 59	56 / 35	40 / 50

Table 3: Average benefit in IBN/PBN conditions for different player dependencies (significant difference in bold)

shown by the table, the combined average benefit for players in all games was 72, which is 10 points greater in the IBN condition than in the PBN condition ($SE = 3$, paired test $t(64) = 1.60, p = 0.04$). Note that in total, there were more goal solicitors than goal revealers, and thus the numbers in the “combined” row do not necessarily equal the summation of “revealing” and “soliciting” players.

Table 1 shows that in both conditions, the benefit to goal solicitors is significantly less than the benefit to goal revealers. It also shows that revealers in the IBN condition increased their benefit by 18 points, as compared to the corresponding games in the PBN condition ($SE = 3, t(29) = 1.7, p = 0.04$). However, there was no significant difference in the benefit between those players that solicited goals in the IBN condition (8 points) and the corresponding games in the PBN condition (9 points). Solicitors in the IBN condition did not lose more benefit than did the corresponding players in the PBN condition. We hypothesized that seeing the revealers’ goals allowed solicitors to offer revealers the chips they needed, in return for chips they did not need. The corresponding player in the PBN condition could not make this kind of offer.

Table 2 presents a pairwise comparison between the number of agreements in both conditions. As can be seen in the table, 16 of the games that resulted in agreement in the IBN condition had failed in the PBN condition. In contrast, only 7 of the games that succeeded in the PBN condition failed to reach agreement in the IBN condition, and this difference was statistically significant ($\chi^2(1, N = 65) = 3.92, p = 0.04$).

The Effect of Players’ Dependency Relationships

Table 3 shows the benefit to players as a function of their dependencies, as well as the number of goals they revealed. As shown by the table, the benefit for dependent players was consistently higher than the benefit for independent players, across conditions. In particular, in those games in the IBN condition where at least one goal was revealed (and the corresponding games in the PBN conditions) they were willing to make a sacrifice. This generosity is considerably more

effective in the IBN condition, where the average benefit to dependent players is 19 points greater than in the PBN condition ($SE = 2.3, t(26) = 2, p = 0.02$). Despite the sacrifice incurred by independent players when goals were revealed, their overall benefit in the IBN condition is 15 points, 17 points greater than the benefit of players in the PBN condition ($SE = 2.3, t(48) = 2.3, p = 0.01$). This is primarily due to their significant gain of 19 points over their PBN scores in IBN games where no goals were revealed.

Lastly, we note that there were significantly more dependent revealers than independent revealers (27 vs. 8), and more independent solicitor than dependent solicitor (24 vs. 11). This explains the finding of Table 1 that the benefit of solicitors is significantly less than the benefit of revealers. Solicitors are likely to be independent players, and assist those dependent players that reveal their goals.

Table 4 shows the number of games that resulted in agreement as a function of the dependency relationship between players. As can be shown by the table, in both conditions, co-dependent players were significantly more likely to reach agreement than dependent players (Fisher’s exact test $p < 0.01$). We also found that the combined performance for co-dependent games was higher than that of single-dependent games, but this difference was not significant, perhaps because of the low number of co-dependent games. In all, there

	single-dep	co-dep	total
IBN	33 (64%)	14 (100%)	47
PBN	26 (50%)	12 (85%)	38

Table 4: Agreement Frequencies

were 9 more acceptances in the IBN condition than there were in the PBN condition (47 compared to 38), and this difference was statistically significant (paired t-test, $p < 0.05$).

Analysis of the IBN Protocol

We now report results relating solely to a set of 112 games played in the IBN condition. This includes additional games that were not played in the IBN condition due to time constraints. Most IBN games did not feature goal revelations, but players revealed their goals the majority (76%) of times they were actually asked. There were 39 games in which one goal was revealed, and 10 games in which two goals were revealed, making for a total of 59 revelations. In all, at least one goal was revealed in 43% of the games. The majority of goal revelations (72%) were performed by dependent players, significantly higher than the number of revelations performed by independent players ($\chi^2(1) = 6.1, p = 0.01$). This confirms the trend shown for the smaller set of games in the previous section.

We also found that it is worthwhile for players to reveal goals: The average benefit for revealers was 64 when agreement was reached, 24 points higher than the benefit in games where agreement was reached without revelation

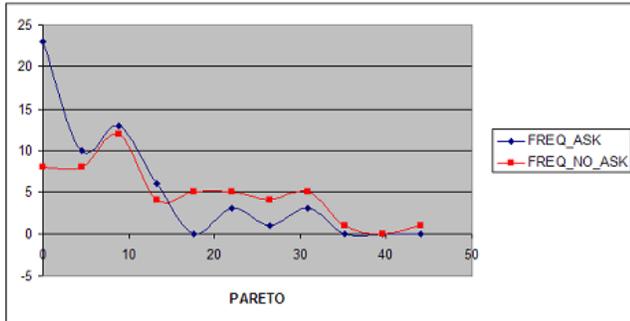


Figure 3: Distribution of games featuring goal solicitations for different numbers of Pareto optimal outcomes possible

($SE = 4.3, t(146) = -2.32, p = .01$). As can be expected, this trend was reversed for goal solicitors, whose benefit in the case where goals were revealed was 22, about 30 points lower than in the case when they were not revealed ($SE = 5, t(146) = -2.8, p = .02$).

We conclude with a summary of goal solicitation behavior against the availability of Pareto optimal outcomes⁴. As shown by Figure 3, a low number of Pareto optimal offers possible in a game results in a greater likelihood that solicitation will occur at some point in that game. However, in games that feature higher numbers of Pareto optimal offers, the trend is reversed, and players are less likely to solicit the goal location of their opponents. This would imply that human negotiators' use of goal solicitation is not random, but is influenced at least partially by the difficulty of the negotiation, the estimation of which correlates with the number of Pareto optimal possible offers.

Discussion and Conclusion

The results shown in the last section establish the role of interest-based negotiation protocols as a mechanism of cooperation in which independent parties solicit dependent parties to reveal information for the purpose of benefiting them. We saw that dependent players are likely to agree to reveal their goals once asked, and that this information is not abused by solicitors. Indeed, they choose to use this information as a tool for assisting the revealers, while succeeding not to incur a loss themselves. This results in a net gain to the revealer, as well as an increase the social benefit of both participants. Lastly, this mechanism is not overused by participants. Solicitors generally dislike to ask for others' goals, and choose to do so mainly in cases where there are few avenues open for beneficial exchanges.

In future work, we plan to conduct further experiments in which players engage in position-based negotiation, but can observe each other's goals. We mean to compare players' performance in those games where goal revelations were solicited in the IBN condition, to see whether the solicitation

⁴An offer is Pareto optimal if no other offer exists that is more beneficial to at least one of the players.

itself (as opposed to simply viewing each others' goals) had an effect on players' performance. Also, we intend to use this data to build better automated negotiation agents based on a sound understanding of how humans negotiate using such tools.

Acknowledgement

Development and dissemination of the Colored Trails formalism is supported in part by the National Science Foundation under Grant No. CNS-0453923. Thanks to Mike Dombrowski for his instrumental work on the CT code base.

References

- Atran, S., Axelrod, R., & Davis, R. (2007). Sacred barriers to conflict resolution. *Science*, (317)(5841), 1039-1040.
- Gal, Y., Grosz, B., Shieber, S., Pfeffer, A., & Allain, A. (2007). The effects of task contexts on the decision-making of people and computers. In *Proc. of the Sixth International Interdisciplinary Conference on Modeling and using Context*.
- Gal, Y., Pfeffer, A., Marzo, F., & Grosz, B. (2004). Learning social preferences in games. In *Proc. 19th National Conference on Artificial Intelligence (AAAI)*.
- Heiskanen, H., Ehtamo, H., & Hamalainen, R. (2001). Constraint proposal method for computing pareto solutions in multi-party negotiations. *European Journal of Operations Research*, 133(1), 44-61.
- Kakas, A., & Pavlos, M. (2006). Adaptive agent negotiation via argumentation. In *Proc. 5th International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS)*.
- Kraus, S., Sycara, K., & Evenchik, A. (1998). Reaching agreements through argumentation: A logical model and implementation. *Artificial Intelligence*, 104, 1-69.
- Loewenstein, G., Bazerman, M., & Thompson, L. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, (57)(3), 426-441.
- Loewenstein, J., & Brett, J. (2007). Goal framing predicts strategy revision: When and why negotiators reach integrative agreements. In *Proc. of the Cognitive Science Society*.
- Rahwan, I., Pasquier, P., Sonenberg, S., & Dignum, F. (2007). On the benefits of exploiting underlying goals in argumentation-based negotiation. In *Proceedings of the 22nd conference on Artificial Intelligence*.
- Rahwan, I., Ramchurn, S. D., Jennings, N. R., McBurney, P., Parsons, S., & Sonenberg, L. (2003). Argumentation-based negotiation. *The Knowledge Engineering Review*, 18(4), 343-375.
- Rubinstein, A. (1985). A bargaining model with incomplete information about preferences. *Econometrica*, 53(5), 1151-1172.
- Vorauer, J., & Claude, S.-D. (1998). Perceived versus actual transparency of goals in negotiation. *Personality and Social Psychology Bulletin*, 24(4), 371-385.