Explaining Negotiation: Obtaining a Shared Mental Model of Preferences

Iris van de Kieft, Catholijn M. Jonker, and M. Birna van Riemsdijk

Man Machine Interaction Group, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

Abstract. Negotiation support systems (NSSs) aim to assist people during the complex process of negotiation. We argue that having a shared mental model of the negotiation task enables and enhances the collaboration between the human negotiator and the NSS. This paper presents an analysis of negotiation that results in a set of concepts that a shared mental model of the user and the NSS should contain. Discrepancies between the individual mental models can arise for various reasons, such as the constructive nature of preferences. Explanation can increase user understanding of the NSS's reasoning, allowing the user to detect and resolve discrepancies. We therefore propose using explanation to achieve and maintain sharedness. We present a framework that provides a means to generate content for such explanations, where we focus on the mental models of user and opponent preferences.

1 Introduction

Negotiation is an interactive decision-making process between two or more parties. It is a complex process that involves emotions as well as computational complexity. As a result, even experienced human negotiators can fail to achieve efficient outcomes [11]. This has motivated the development of *negotiation support systems* (NSSs). These software systems assist a human negotiator (user) in negotiation by, for example, aiding communication, enhancing negotiation skills, and reducing cognitive task load.

The Pocket Negotiator project, see [4], strives for synergy between NSS and the human negotiator. The NSS and the user should work together as a *team* in which their complementary skills are needed to achieve good outcomes. It is well-known from the social psychology literature that performance of human teams is positively influenced by the team members having a shared understanding or *shared mental model* of the task and the team work involved ([5,7]). The concept of shared mental model is defined in [3] as:

> knowledge structures held by members of a team that enable them to form accurate explanations and expectations for the task, and, in turn, coordinate their actions and adapt their behavior to demands of the task and other team members.

K.G. Mehrotra et al. (Eds.): IEA/AIE 2011, Part II, LNAI 6704, pp. 120–129, 2011.
© Springer-Verlag Berlin Heidelberg 2011

We maintain that having a shared mental model is not only important in human teams, but also in human-agent teams. The representation that an automated agent has of a task can be viewed as its mental model. Discrepancies between the mental models of the NSS and the user may at best result in innocent misunderstandings but at its worst may result in a dysfunctional cooperation.

This paper contributes to the technology to achieve sharedness between mental models of NSS and user. In Sect. 2, we first analyze negotiation and the interaction between user and NSS, to determine the essential components of such a shared mental model . Furthermore, the analysis reveals the possible causes of discrepancies between mental models, for example, the constructive nature of preferences.

Based on our analysis, in Sect. 3 we identify *explanation* (see for example [12,13]), as a suitable technology for improving sharedness between mental models. Explanations can increase the transparency of the system, allowing the user to detect and resolve discrepancies. We provide a structured approach to generating such explanations. Due to space restrictions, we focus on one part of the shared mental model: user and opponent preferences. We describe different levels of content selection, for explanations of bids and their utilities. We leave the form (i.e., presentation) of the explanation for future work.

2 Bilateral Multi-issue Negotiation

Bilateral multi-issue negotiation is the process in which two parties try to reach an agreement about multiple issues. The following four major stages can be discerned in integrative negotiation: private preparation, joint exploration, bidding, and closing. *Private preparation* is about information gathering and reflection before meeting the other party. In *joint exploration* the negotiating parties talk to each other, but do not place bids on the table. During *bidding*, both negotiators exchange bids according to the agreed protocol, typically a turn-taking protocol. During the *closing* stage the outcome of the bidding stage is formalized and confirmed by both parties.

We first present a basic negotiation framework, followed by an analysis of the negotiation process. This analysis discusses human weaknesses in negotiation and subsequently focuses on the interaction between user and NSS.

2.1 A Basic Negotiation Framework

Let $p \in \{s, o\}$ represent the negotiating parties "Self" (s) and "Opponent" (o). Let I, with typical element i, denote the finite set of issues under negotiation. For example, in a job negotiation I might consist of the issues *salary, car, vacation days*. For convenience we assume issues to be numbered from 1 to n, where $n = |I|$, and henceforth, we will refer to issues by their respective numbers. For example, issue 1 is salary, issue 2 is car, issue 3 is vacation days.

For each $i \in I$ let V_i be the set of possible values that this issue may have. For example, the issue *car* may have {yes, no} as possible values. Let $V = V_1 \times V_2 \times \ldots \times V_n$, be the set of possible outcomes. A possible outcome $v \in V$ is thus an n-tuple $\langle v_1, v_2, \ldots, v_n \rangle$. A possible outcome for the job domain can, for example, be $\langle 2000, no, 20 \rangle$.

A negotiation domain is denoted by $D = \langle I, V \rangle$. A bid in domain D is denoted by b^p, with $b \in V$ and p the party that proposed the bid. For simplicity, a bid may be denoted as b, if p is understood or not relevant.

Each party p has a utility function $u_p : V \rightarrow [0, 1]$ which assigns a utility between 0 and 1 to possible outcomes and bids. A commonly used type of utility function is the linear additive normalized function, defined by $u^p(b) = \sum_{i=1}^{n} u_p^i(b)$, where $u_p^i(b)$ is the utility of issue value b_i for party p. That function is defined in terms of a weight w_p^i and evaluation function e_p^i for that issue and party: $u_p^i(b) = w_p^i e_p^i(b_i)$. The weight is the relative importance that party p assigns to issue i in such a way that $\sum_{i=1}^{n} w_p^i = 1$, for all $p \in \{s, o\}$. The evaluation function $e_p^i : V_i \rightarrow [0, 1]$ assigns a score between 0 and 1 to the possible values of issue i. We use the notation $\vec{u}_p(b)$ to represent the n-tuple $\langle u_p^1(b), u_p^2(b), \ldots, u_p^n(b) \rangle$. Similarly, $\vec{e}_p(b) = \langle e_p^1(b_1), \ldots, e_p^n(b_n) \rangle$ and $\vec{w}^p = \langle w_p^1, \ldots, w_p^n \rangle$.

2.2 The Weaknesses of the Human Negotiator

In this first part of our analysis, we discuss the problems humans have with negotiation, assuming there is no NSS support. There are two ways to categorize the problems humans have with negotiation: related to *outcome*, or related to the negotiation *process*. The outcome related pitfalls in negotiation are: leaving money on the table, settling for too little, rejecting a better offer than any other available option, and settling for terms worse than alternative options [1,11].

The outcome related pitfalls are caused by the problems people have during the negotiation process, which are related to the following (see [1,11] for more information):

- *Lack of training:* Without training, humans often have difficulty in structuring negotiation problems and thinking creatively about such problems.
- *Lack of preparation:* Preparation is insufficient when it leaves the negotiator unaware of an important part of the issues, underlying interests, the preferences and/or circumstances of the parties involved, see for example,.
- *Structural barriers to agreement:* This refers to such problems as die-hard bargainers, a bad atmosphere, power imbalance [6], cultural and gender differences, disruptive or incommunicative people, and a lack of information.
- *Mental errors:* Parties commit mental errors such as the escalation error, biased perception, irrational expectations, overconfidence, and unchecked emotions.
- *Satisficing:* Due to uncertainty of the future, the costs of acquiring information, and the limitations of their computational capacities, people have only bounded rationality, forcing them to make decisions by satisficing, not by maximization.

This difficulties indicate why it may be difficult for a human to have an accurate mental model of a negotiation.

2.3 The Interaction between Human Negotiator and NSS: What to Share?

In this section we analyze the interaction between user and NSS, to provide insight into their task division. This analysis helps determine the contents of the shared mental model that needs to be cultivated between user and NSS.

Negotiation is a prime example of a task for which the human mind is only partially equipped, and for which artificial intelligence can only provide partial assistance. For example, the user has a wealth of knowledge about the world and about interacting with other humans, but need not be a specialist in negotiation. The NSS specializes in negotiation. It makes generic negotiation knowledge available to the human. Also, the user has limited working memory and limited computational power, i.e., bounded rationality. The NSS has better memory and can search much more quickly through much larger outcome spaces. This implies that tasks should be divided between user and NSS in a way that respects their complementary capabilities.

This task division suggests they need not share all their knowledge. However, some shared information is necessary for cooperation, hence the need for a shared mental model. The information and knowledge exchange between these two team members is as follows: during the preparation and exploration stage the user needs to inform the NSS about the current negotiation, e.g., the Opponent, the set of issues I, and outcome space V, and the utility functions of Self and Opponent. These utility functions are a model of the actual preferences of Self and Opponent. The actual preferences may not be fully known, and may be subject to change. The NSS needs this user input in order to provide assistance during the bidding stage, when strategic bidding decisions have to be made.

For this information exchange to be successful, the user must fully understand the process of negotiation and what is expected of him/her by the NSS, and what can be expected in return. This implies that during the negotiation stages, the NSS needs to provide the user (upon request) with generic negotiation information, but also current negotiation information regarding the Opponent, I, V, and utility functions, in as far as such information is available to the NSS.

Thus, a shared mental model of a human negotiator and an NSS should at least contain submodels on:

- domain knowledge D
  - I: set of issues
  - $\forall i \in I$: V_i the value range of issue i
- knowledge about negotiating parties Self and Opponent. For each $p \in \{s, o\}$:
  - u_p: the utility function of p, in so far as known
  - the emotional status and coping style of p
  - the negotiation model of p
- knowledge about the capabilities of the team members: Self and NSS

– bidding knowledge
  - bidding history: the sequence of bids that have been exchanged so far
  - the current bidding strategy for Self
  - the bidding protocol, including information about available time

2.4 Discrepancies between Mental Models

Based on the previous subsections, we identify what may cause lack of sharedness with respect to the elements of the shared mental model, or in other words, what may cause *discrepancies* between the mental models of the user and the NSS. We consider a discrepancy between mental models to exist when one model contains information regarding an element, and the other model contains either conflicting information regarding this element, or no information regarding this element. Once a discrepancy is detected, it can be resolved by adapting (one of) the mental models.

One particular aspect that may lead to discrepancies is the *constructiveness* of domain and preference information. Even with proper preparation, information on the domain and preferences of Self and Opponent is often difficult to determine fully at the start of the negotiation. Humans have been found to discover this information along the way. Due to this constructiveness, the user may discover new knowledge during the negotiation that the NSS does not yet have, thus causing a discrepancy. Table 1 lists, for each team member, some possible causes for their mental model to lack (correct) information.

Table 1. Causes for lack of (correct) information in mental models

User mental model		**NSS mental model**
-lack of training	-lack of preparation	-lack of user input
-constructive domain	-bounded rationality	-constructive domain
-constructive preferences of Self and Other		-constructive preferences of Self and Other

3 Explanation Framework: Resolving Discrepancies

Explanation can serve various purposes, such as improving effectiveness (helping users make good decisions), increasing the users trust in the system and improving *transparency* of the system [12]. In this paper we are in particular interested in the latter, as this facilitates detection and resolving of discrepancies between mental models of NSS and user. Transparency means explaining how the system works, thus giving the user a better understanding of the NSS's reasoning process. This allows the user to detect any discrepancies between the mental models, and subsequently to resolve these discrepancies by updating the mental models where necessary.

Due to space restrictions, for the remainder of this paper, we focus on the mental models of preferences, as represented by the utility functions u_S and u_O. The different causes of discrepancies between the mental models (of preferences) suggest that different types of preference information need to be made transparent in different situations. First, lack of training can result in difficulty in switching between one's own perspective and that of the Opponent. The user may neglect the preferences of the Opponent, whereas the NSS often uses this information for bid calculations. This suggests that *both user preferences and opponent preferences* should be presented to the user, so that the user remains aware of both perspectives. This will make it easier for the user to understand the NSS's bid recommendations.

Second, due to bounded rationality, users may have difficulty calculating bid utilities. To make transparent why bids are assigned a certain utility value, *more insight into the utility function* needs to be available. If the NSS assigns a utility that the user believes is too high, it is not always immediately clear to the user why it is too high. In order to resolve the discrepancy, the user needs to be able to determine where the problem lies exactly (i.e., in which part of which mental model).

As we assume the utility function is a linear additive function, we distinguish three levels of insight into this function. The first level is the overall utility, $u_p(b)$. The second level concerns the building blocks of the first level, the issue utilities: $u_p^i(b)$. The third level concerns the building blocks of the second level: the weights and evaluation functions of the issues: w_p^i and $e_p^i(b_i)$.

Third, also due to bounded rationality, users may have difficulty calculating *utility differences between bids*. This suggests that these differences should also be made transparent. The NSS may determine that bid b is better than bid b', however, the user may also want to know how much better. Explanation about utility differences reduces the chance of mental errors such as irrational expectations.

In this section, we present a framework that can be used as the basis for generating explanations to increase transparency with respect to the NSS's mental model of preferences. Research has distinguished between the content and form of explanations (e.g., [9]). The focus of the proposed framework is content selection; how it should be presented is an additional step that we only touch upon lightly in this paper. Further work will address this in more detail.

The explanation content is selected from the so-called *originator* [10], which in our case is the original NSS without explanation capabilities. The content is selected by the so-called *explainer* [10], a component that is to be added to the NSS. The explainer is in charge of generating explanations. Given a bid b, we assume that the utility functions $u_p(b)$ and their sub-parts $u_p^i(b)$, w_p^i and $e_p^i(b_i)$ are provided by the originator. When comparing two bids, additional content regarding differences in utility is necessary. This may not be readily available in the originator, we thus leave it to the explainer to perform these calculations. Any additional knowledge that should be present in the explainer will be described at the relevant points in the framework.

Content selection is presented for each of the three levels of detail of the utility function described above. For each level, the framework provides the content to be selected when *evaluating a single bid*, as well as the additional content needed when *comparing two bids*.

3.1 Overall Utility Level

This level concerns the overall utility function: $u_p(b)$. When evaluating a *single bid* b at this level, the content selection, referred to as CS1, consists of the bid itself and the overall utility $u_p(b)$ for each p involved:

$$\text{CS1}(b) = \langle b, u_s(b), u_o(b) \rangle$$

When *comparing two bids* b and b', we first select the relevant content for a single bid (CS1). Additionally, the explainer calculates the difference in overall bid utility for each party. The difference in overall utility for party p between two bids b and b' is defined as follows: $\Delta u_p(b, b') = u_p(b') - u_p(b)$. Then, the content selection, referred to as CS2, is:

$$\text{CS2}(b, b') = \langle \text{CS1}(b),\ \text{CS1}(b'),\ \Delta u_s(b, b'), \Delta u_o(b, b') \rangle$$

This content could, for example, be presented to the user as follows: "Bid b' has $\Delta u_s(b, b')$ more utility for you than bid b. For your opponent bid b' has $\Delta u_o(b, b')$ more utility than bid b." A graphical display could supplement this text, listing the utilities per bid per party, such as in Table 2.

Table 2. Content for overall utility level

bid	**utility self**	**utility other**
b	$u_s(b)$	$u_o(b)$
b'	$u_s(b')$	$u_o(b')$
	$\Delta u_s(b, b')$	$\Delta u_o(b, b')$

3.2 Issue Utility Level

This level concerns the utility functions per issue, $u_p^i(b)$. Here, more detail is provided as to how the utility of the bid was calculated, by showing how the utilities per issue together determine the overall utility. For a *single bid* b, we select the bid and the n-tuple of the utility values per issue, for each party:

$$\text{CS1}(b) = \langle b, \vec{u}_s(b), \vec{u}_o(b) \rangle$$

Furthermore, additional explanation knowledge is needed, to explain to the user how the utilities per issue are combined to get the overall utility. Because here we use an additive utility function, the additional content consists of the fact that the combination of the issue utilities is additive.

When *comparing two bids* b, b' at this level, we select the relevant content for each bid separately (CS1). Additionally, we provide content to show which issues have different values, and what the associated difference in utility is. We first define the difference in utility of issue i for party p, between bid b and b':

$\Delta u_p^i(b, b') = u_p^i(b') - u_p^i(b)$. The corresponding n-tuple of utility difference for each issue is: $\vec{\Delta u}_p(b, b')$ The content selection is then defined as follows:

$$\text{CS2}(b, b') = \langle \text{CS1}(b),\ \text{CS1}(b'),\ \vec{\Delta u}_s(b, b'), \vec{\Delta u}_o(b, b') \rangle$$

As this content consists of several n-tuples, a textual presentation does not seem suitable; a graphical presentation such as Fig. 1 seems more appropriate. In this figure, the utility differences are not presented as such, however, this information is used to highlight certain utility changes. In general, focus can be created by highlighting just the issues that have different values (*DiffIssues*), increased utility value (*PosDiff*), decreased utility value (*NegDiff*), or the same utility value (*NoDiff*). This could be made into a dynamic feature for the user to interact with. We therefore define the following sets of issues:

$$\begin{aligned} \mathit{DiffIssues}(b, b') &= \{i | b_i \neq b'_i\} \\ \mathit{PosDiff}^p(b, b') &= \{i | \Delta u_p^i(b, b') > 0\} \\ \mathit{NegDiff}^p(b, b') &= \{i | \Delta u_p^i(b, b') < 0\} \\ \mathit{NoDiff}^p(b, b') &= \{i | \Delta u_p^i(b, b') = 0\} \end{aligned}$$

In addition to highlighting purposes, these sets can be used to select a subset of the content. For example, if only one issue differs between bids b and b' (i.e., $|\mathit{DiffIssues}(b, b')| = 1$), we might choose to only present the content related to that issue $i \in \mathit{DiffIssues}(b, b')$.

	Bid b			Bid b'		
issue	*value*	*utility Self*	*utility Other*	*value*	*utility Self*	*utility Other*
salary	**2000**	0.3	0.4	**3000**	0.4	0.3
car	**yes**	0.2	0.1	**yes**	0.2	0.1
vacation	**24**	0.3	0.4	**20**	0.25	0.4

Fig. 1. Example presentation of explanation content for two bids, two parties, at issue utility level. The utilities highlighted in white indicate a decrease in utility, the utility highlighted in orange indicates an increase in utility, when comparing b and b'

3.3 Level of Issue Weights and Evaluation Functions

This third level provides more detail as to why an issue has a certain utility value. This concerns the weights and evaluation functions of the issues, w_p^i and e_p^i, which together determine the utility assigned to an issue value. The content when evaluating a *single bid b* consists of the issue weights and the evaluation function score of each issue value:

$$\text{CS1}(b) = \langle b, \vec{w}_p, \vec{e}_p(b) \rangle$$

Additionally, the explanation should contain the information that the weight and evaluation function score are multiplied to give the issue utility.

Several kinds of *comparisons* can be made at this level. For example, comparisons can take place for one issue across two bids, for two issues within one bid or for two issues across two bids. Due to the many possible comparisons, this information should only be presented to the user when requested. We define $\Delta w_p(i,j) = (w_p^j - w_p^i)$ and $\Delta e_p(b_i, b'_j) = e_p^j(b'_j) - e_p^i(b_i)$. The explainer can use these formulas together with CS1 to determine the content necessary for a requested comparison.

An explanation may clarify why an issue i has $\Delta u_s^i(b, b')$ more utility for the user in b' than in b. Alternatively, an explanation may clarify why, for bid b, issue i is assigned more utility than issue j. For example, the explanation may be: "Although issue j has $\Delta w_p(i,j)$ more weight than issue i, issue value b_i has a higher evaluation score $(\Delta e_p(b_i, b_j))$ than issue value b_j. When multiplied, this results in a higher issue utility for value b_i of issue i."

4 Conclusion

We presented an approach in which we analyzed the cooperation between user and NSS from the perspective of shared mental models. This analysis let us determine the elements that should be part of the shared mental model between user and NSS. This analysis also served to determine possible causes of discrepancies between mental models.

Following this analysis, we focused on resolving discrepancies for one part of the shared mental model: preferences. We proposed using explanation, as it is a means to increase user understanding of the reasoning of the NSS, which helps detect and resolve discrepancies. We then presented a framework that provides a formal description of content selection for explanaining preferences. We focused on bilateral multi-issue negotiation, nevertheless, as the general idea of explaining preferences also applies to other types of negotiation, such as multilateral negotiation, our work should be extendable to such other types.

Future work includes investigating how to present the explanation content (i.e., visualization techniques), a procedure to determine when to present which explanation to the user, implementation of the framework and user tests to validate it, and extending this work to other aspects of the shared mental model. Formalizing our approach of analyzing the problem domain in the light of shared mental models would allow it to be applied to other decision support domains.

Also, as integration between explanation and argumentation research has been proposed in [8], it would be interesting to investigate how we might integrate our explanation framework, which focuses on the user-NSS interaction, with argumentation-based negotiation (e.g.,[2]), which is used for the interaction between user and opponent.

Acknowledgements. This research is supported by the Dutch Technology Foundation STW, applied science division of NWO and the Technology

Program of the Ministry of Economic Affairs. It is part of the Pocket Negotiator project with grant number VICI-project 08075.

References

1. Harvard Business Essentials: Negotiation. Harvard Business School Publishing Corporation, Boston (2003)
2. Amgoud, L., Dimopoulos, Y., Moraitis, P.: A unified and general framework for argumentation-based negotiation. In: Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 1–8. ACM Press, New York (2007)
3. Cannon-Bowers, J., Salas, E., Converse, S.: Shared mental models in expert decision making teams. In: Current Issues in Individual and Group Decision Making, pp. 221–246 (1993)
4. Hindriks, K., Jonker, C.: Creating human-machine synergy in negotiation support systems: towards the pocket negotiator. In: Proceedings of the 1st International Working Conference on Human Factors and Computational Models in Negotiation, pp. 47–54. ACM, New York (2008)
5. Lim, B., Klein, K.: Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. Journal of Organizational Behavior 27(4), 403–418 (2006)
6. Mastenbroek, W.: Negotiating as emotion management. Theory, Culture & Society 16(4) (1999)
7. Mathieu, E., Heffner, T., Goodwin, G., Salas, E., Cannon-Bowers, J.: The influence of shared mental models on team process and performance. The Journal of Applied Psychology 85(2), 273–283 (2000)
8. Moulin, B., Irandoust, H., Bélanger, M., Desbordes, G.: Explanation and argumentation capabilities: Towards the creation of more persuasive agents. Artificial Intelligence Review 17(3), 169–222 (2002)
9. Pu, P., Chen, L.: Trust-inspiring explanation interfaces for recommender systems. Knowledge-Based Systems 20(6), 542–556 (2007)
10. Roth-Berghofer, T., Richter, M.: On explanation. Künstliche Intelligenz 22(2), 5–7 (2008)
11. Thompson, L.: The mind and heart of the negotiator. Pearson Prentice Hall, NJ (2005)
12. Tintarev, N., Masthoff, J.: A survey of explanations in recommender systems. In: ICDE 2007 Workshop on Recommender Systems and Intelligent User Interfaces, pp. 801–810 (2007)
13. Ye, L., Johnson, P.: The impact of explanation facilities on user acceptance of expert systems advice. MIS Quarterly 157–172 (1995)