Endogenous Preferences in Multi-Issue Bargaining

Henner Gimpel*

*Information Management and Systems, University of Karlsruhe
Englerstr. 14, 76131 Karlsruhe, Germany
http://www.iw.uni-karlsruhe.de

gimpel@iw.uni-karlsruhe.de

Abstract

Negotiation analysis and game theoretic bargaining models usually assume parties to have exogenous preferences from the beginning of a negotiation on and independent of the history of offers made. On the contrary, this paper argues that preferences might be based on issue-wise reference points changing during the bargaining process. Endogenous preferences are thereby outlined, an internet experiment on the construction of preferences in a bilateral negotiation is presented, and implications for negotiation support systems are discussed.

1 Introduction

In a negotiation, two individuals have interests in reaching one of several possible agreements, but their preferences over these agreements are not completely identical. In multi-attribute negotiations, studied here, parties usually have the possibility to simultaneously negotiate over several attributes and to search for integrative potential—they play a non-constant-sum game.

Empirically, negotiators oftentimes fail to reach mutually beneficial agreements in integrative multi-issue negotiations and—even if they reach an agreement at all—it is oftentimes non-Pareto optimal [5, 14, 15, 25]. It appears that there is a serious behavioral bias preventing Pareto-optimal solutions, namely the *endowment effect* leading to endogenous preferences changing over the course of the bargaining process. The questions at hand are:

1. Do preferences change during a negotiation upon receiving an offer?
2. If so, how do negotiators update their preferences?
3. What are the implications for the design of a negotiation support system?

These questions are examined in the following. To this end, the remainder of the paper is structured as follows: at first, the notion of endogenous preferences is outlined in Section 2. Section 3 then displays graphically how negotiators might update their preferences during a negotiation and Section 4 reports on an internet experiment testing for endogenous preferences. Finally, Section 5 surveys related work and Section 6 concludes with pointing out implications for negotiation support systems.

2 Endogenous Preferences

Integrative bargaining is oftentimes regarded as joint problem solving [4] and jointly exploring the space of potential agreements; search techniques like logrolling and trade-off heuristics for finding mutually beneficial agreements are proposed [8, 21]. However, many practical heuristics and advices how to negotiate have the limitation of assuming preferences to be exogenously given and fix during the bargaining process[1]—this may not be the case in many real life negotiations.

The value of a good to an individual increases when the good becomes part of the individual's endowment [9]. This effect—termed *endowment effect* [22]—grows stronger with duration of ownership and prevails after termination of ownership [20]. During a negotiation process, parties constantly gain *virtual possession* on single issues and afterwards feel entitled to the specific value on this issue. Negotiators value the counterparty's proposals relative to a reference point and this reference point is adjusted issue-wise during the bargaining process. A proposal which is seen as a loss relative to the current reference point on one or more issues, is devalued—preferences are endogenously changing and depend on the history of the specific negotiation. Each single offer proposed to the counterparty might change the counterparty's preferences and therefore might destroy potential agreements. Searching the agreement space extensively becomes prohibitive.

Cray and Kersten report on data gathered with the Inspire system [5]: in a pre-negotiation phase they elicit preferences, then they conduct the negotiation, and—if an inefficient agreement is reached—they present Pareto-improvements to the negotiators. 60.2% of the agreements in their data set were inefficient with respect to the preferences elicited in the first phase. However, only 20.8% of negotiators were willing to accept the proposed Pareto-improvements. At first sight this seems puzzling and irrational; but if preferences are endogenous and change during the negotiation, the system's proposal in the post-settlement phase may be unacceptable with respect to the ex-post preferences. This might explain the low acceptance rate.[2]

Vetschera analyzes utility functions, offers made, and final agreements in thousands of Inspire negotiations [25]. He reports that in about 25% of the cases, negotiators violated consistency in the sense that their observed behavior did not fit the ex-ante elicited utility functions. Negotiators' inconsistency could possibly be tracked back to a changed preference

[1] Furthermore, preferences are usually assumed to be unknown to the counterparty.

[2] Other explanations might be that Inspire presents several possible Pareto-improvements and the parties would have to enter a new negotiation for agreeing on one of them. This new negotiation is afflicted with costs (e.g. mental effort and time) and might therefore not be desirable for negotiators. Furthermore, parties might tend to ex-post rationalize their behavior after they negotiated an agreements; they might stick to the agreement for not admitting that they could have done better.

structure by the time the seemingly inconsistent behavior is observed.[3]

Endogenous reference-dependent preferences and the process of issue-wise updating reference points is exemplified in the following.

3 Graphical Example

This section exemplifies reference points endogenously changing issue-wise during a two-issue alternating-offer negotiation involving negotiators A (she) and B (he). The process is displayed in a sequence of Edgeworth-boxes[4] in Figure 1.

Both issues, named x and y, are normalized to one and both parties prefer more to less on both issues. The share of negotiator A is measured from the boxes lower left corner. Accordingly, negotiator B's share is measured from the upper right corner. For the sake of simplicity, only negotiator A has a shifting reference point in this example.

Figure 1a shows the initial condition at time $t = 0$; time is indicated by superscripts. Three of negotiator A's indifference curves are displayed and the associated utility levels are indicated at the top of the figure; $u^0_{A,3} > u^0_{A,2} > u^0_{A,1}$ holds in this example. Furthermore, A has a reference point r^0_A which might initially be at zero on both issues and A makes an offer o^0_A at $t = 0$. Here, A is quite greedy and demands the best possible solution for herself.

At $t = 1$ (1b), B makes a counteroffer anywhere in the interior of the Edgeworth box. It is named o^1_B.

At $t = 2$ (1c), it is A's term again and the endowment effect applies the first time: A sees what she is offered, namely o^1_B, and feels entitled to this level of x and y. Not, that it would already be hers—after all the negotiation is still going on—but she now feels a *virtual endowment* and might assume that she never has to accept less on any issue. A's reference point shifts to r^2_A which is identical with o^1_B here. Furthermore, A makes the new offer o^2_A.

Three things about shifting the reference points are noteworthy here: Firstly, r^2_A equals o^1_B in this example. This is not necessarily the case. If one thinks about the update function f in the single-issue model of shifting reference points presented by Compte and Jehiel [3] which is discussed in Section 5, then the reference point could be anywhere between r^0_A and o^1_B. Such a partial adaption would be in line with the finding that reference point adaption is not all or nothing but a rather slowly progressing process, as presented by Strahilevitz and Loewenstein [20]. Secondly, it is not only that A expects to get a utility of at least $u^0_{A,1}$. A expects to get at least as much on x, as r^2_A is in this attribute and at least as much on y, as r^2_A gives. Thirdly, this process of shifting the reference point will go over the entire negotiation and there will be a history of (virtual) ownership effect (cf. [20]).

At $t = 3$ (1d), B rejects o^2_A and counters with o^3_B. Subsequently this is rejected by A at $t = 4$ (1e). The negotiation goes on, but the interesting point is reached here, so the description

[3] Again, endogenous preference changes are not the only possible explanation for the behavior. Nuisance in the specific utility elicitation technique employed in the first phase of negotiation support might, for example, serve as a reason for the observed inconsistencies as well.

[4] Called *Pareto-boxes* by some authors.

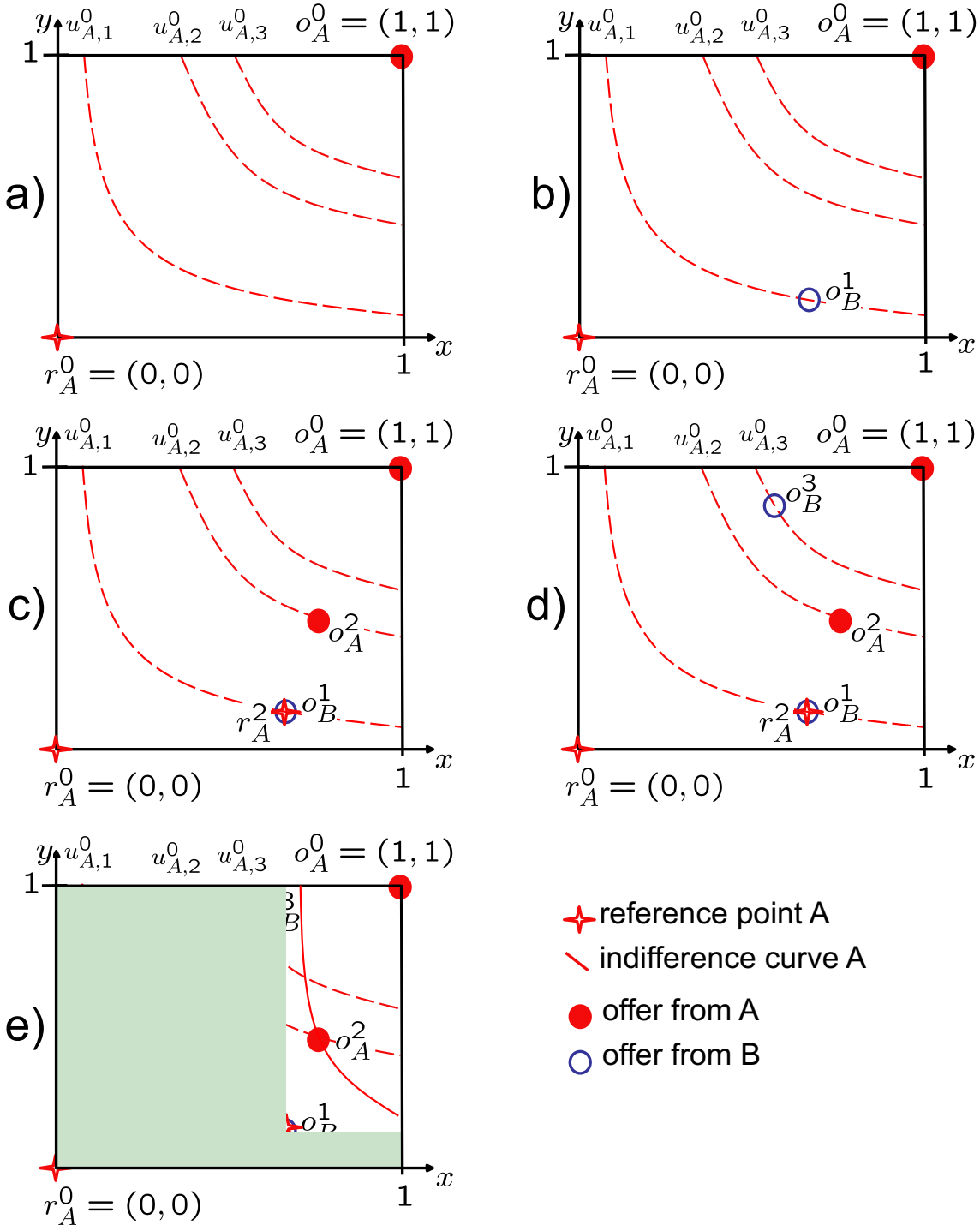

Figure 1: Bilateral Negotiation in an Edgeworth box

ends.

The rejection of o_B^3 is somewhat puzzling. With respect to A's preferences at $t = 0$, given by the dotted indifference curves, o_B^3 gives utility $u_{A,3}^0$ which is strictly more than $u_{A,2}^0$. So B offers A a higher utility at $t = 3$, than A asked for at $t = 2$. However, A's preferences from $t = 0$ might no longer be the current preferences if A evaluates o_B^3 with respect to r_A^2. On attribute y, o_B^3 is much better for A, then the reference point; on the other hand, it is a little worth on x. Every potential agreement worse than r_A^2 on any issue is shaded in Figure 1e.

If A would be concerned with (ex-ante) utility only, she would accept o_B^3, as the huge increase in y over-compensates the small loss in x. But if she cares about reference points and issue-wise losses or gains, she (irrationally ?) rejects o_B^3. The interesting question now is, whether such patterns and evidence for endogenously changing preferences can be found in empirical data.

4 Experiment

The notion of endogenous preferences was tested in an internet experiment (see [16] and [17] for an introduction to internet-based experimentation). The main challenge for testing on endogenous preferences in an experiment is, that subjects' preferences are neither directly observable, nor can they be elicited reliably—at least not multiple times from the same subject.

The basic idea of the experiment is to confront subjects with different negotiation strategies in different treatments and to measure their ex-post preferences. These ex-post preferences can then be analyzed in a between-subject comparison. If assignment of subjects to treatments is randomized, there should be no systematic differences in the subjects ex-post preferences across treatments except if the preferences are influenced by the process of negotiating. Systematic differences across treatments thus support the hypothesis that preferences are endogenously formed in a negotiation.

The operationalization of the basic idea is as follows: subjects negotiate bilaterally with a software agent on a tenancy contract. The monthly rent, the availability of an elevator, and the existence of a balcony are negotiable attributes in the tenancy contract. All other attributes like the size of the apartment, the available furniture, the location, etc. are non-negotiable and fixed in the subjects' instructions. After negotiating, the subjects willingness to accept a change in a single attribute of the negotiated agreement is elicited and compared across treatments.

The experiment took place in April 2005 at the University of Karlsruhe, Germany. The procedure was as follows: subjects' were recruited in an undergraduate class on business administration; the lecturer briefly promoted participation and sheets of paper with the login data were handed out. Subjects then logged in from home over the internet via a web browser. A subject's session started with two pages of instructions followed by a questionnaire. Afterwards the subject negotiated over a tenancy contract. The strategies will be described afterwards. After the negotiation terminated, a second landlord stepped in and offered an alternative contract. This second landlord served for giving subjects in different treatments the same final tenancy contract. From this contract, the subject's willingness to accept a wors-

ening in either the attribute elevator (attribute A) or the attribute balcony (attribute B) was elicited by asking the subject to solve to indifference equations for the respective monthly rent, i.e. the monthly rent was taken as numeraire. Finally, the session ended with retrieving some demographic data, thanking the subject, and asking for general comments on the experiment. The entire procedure was explained to subjects in the experiment's instructions. A week after recruiting the subjects, 50 Euro were awarded to one of the experiment's participants in a lottery held in the lecture. Each participant had the same chance of winning the lottery which was not related to a subject's specific choices during the experiment. Therefore, the lottery served as an incentive for participation but was not a salient reward as oftentimes used in experimental economics (cf. [19] and [7, p. 24ff]).[5]

During the negotiation, offers for contracts are exchanged between the two parties. The software agent representing the landlord starts with an initial offer and subsequently the parties alternate in deciding on the acceptability of an offer and proposing a counteroffer after rejecting an offer. If a maximum of six offers per party is reached, i.e. overall twelve offers, without any of the offers being accepted, an arbitrator steps in and proposes an agreement which is binding for the two parties. The subject can, however, improve its contract by accepting the second landlord's proposal which always is strictly better than the arbitrator's proposal.

A critical point in analyzing the data gathered in the experiment is whether the subjects were sufficiently motivated to participate and sincerely consider the choices presented to them. As the financial reward was non-salient, this might be questioned. However, casual observations indicate that the subjects were motivated. Immediately accepting the landlord's first offer would have been the fastest way to secure participation in the lottery. However, none of the subjects chose this least-effort-way and just one subject accepted the landlord's second order. All other subjects eagerly negotiated. Overall, 47 students logged in at the experimental system; this is about 90% of the students addressed in the lecture. In four cases, the session was abandoned by the respective subject and one observation was discarded as the subject used the web browser's forward and backward functionality.[6] Overall, there are 42 valid observations and the subjects' intrinsic motivation for participating sincerely seems to be rather good.

The experiment embraces four treatments, i.e. four classes of strategies. Strategies in treatment 1 ($T1$) have in common that each single offer grants an elevator to the subject. In treatment 2 ($T2$), no offer grants an elevator to the respective subject. The values of the other attributes are randomized with a slight tendency to making concessions as the negotiation progresses. Treatments 3 and 4 ($T3$ and $T4$) utilize the strategies employed in T1 and T2 with the difference that attributes A and B are reversed, i.e. $T3$ always offers a balcony to the subject while randomizing the values of the remaining two attributes and $T4$ never offers a balcony. The software agents simply presents the sequence of predefined offers and does not accept any subject's offer unless it dominates either a previous offer by the agent or the next offer to be made.

According to a conventional microeconomic model, the willingness to accept should be the same in all four treatments and for both attributes individually. More formally, this means

[5]The full instructions as well as screenshots from the system are available upon request.

[6]In the instructions, subjects were asked not to use the browser's forward and backward functions as this allows them to retract to a previous offer by the agent for accepting it after seeing the agents next offer.

that the equations

$$WTA_A^{T1} = WTA_A^{T2} = WTA_A^{T3} = WTA_A^{T4}$$

and

$$WTA_B^{T1} = WTA_B^{T2} = WTA_B^{T3} = WTA_B^{T4}$$

should hold. WTA thereby stands for the willingness to accept a worsening in the attribute denoted by the subscript in the treatment indicated by the superscript.

On the contrary, a behavioral view on negotiator decision making suggests that subject in $T1$ might feel more entitled to getting an apartment with an elevator than subjects in $T2$ do. This entitlement would lead to a higher WTA on attribute A in treatment $T1$ than in treatment $T2$. For attribute B, i.e. the balcony, there shouldn't be a difference across $T1$ and $T2$. Furthermore, subjects in $T3$ might feel more attachment towards a balcony than subjects in $T4$ while there shouldn't be a difference concerning the elevator. Overall, the behavioral predictions are:

$$WTA_A^{T1} > WTA_A^{T2}$$
$$WTA_B^{T1} = WTA_B^{T2}$$
$$WTA_A^{T3} = WTA_A^{T4}$$

and

$$WTA_B^{T3} > WTA_B^{T4}$$

The average willingness to accept in T1 and T2 is given in Table 1. The different predictions derived from the conventional and the behavioral model can be tested with the data. The conventional model predicts no systematic difference between treatments whereas the behavioral model predicts a difference for attribute A. For inferences, a randomization test on ranks of classified data is used. The difference across treatments with respect to attribute A is significant at a 5% level (one-sided test), whereas the difference with respect to attribute B is not significant at any reasonable level (two-sided test). For attribute B, both models coincide with their prediction and this prediction cannot be rejected. For attribute A however, the conventional prediction can be rejected, whereas the behavioral prediction does a better job at explaining the data.

Table 1: Average WTA in treatments 1 and 2, taking the monthly rent as numeraire

treatment	sample size	Average WTA attribute A	Average WTA attribute B
T1	13	12.9	10.7
T2	14	4.9	14.1

The average willingness to accept in T3 and T4 is given in Table 2. Again, a randomization test on ranks of classified data is used for inferences. The difference across treatments with respect to attribute A is not significant at any reasonable level (two-sided test), whereas here—as suggested by the behavioral model—the difference with respect to attribute B is

significant at the 5% level (one-sided test). For attribute A, both models' predictions coincide and cannot be rejected. For attribute B, the conventional prediction can be rejected, whereas the behavioral prediction again does a better job at explaining the data.

Table 2: Average WTA in treatments 3 and 4, taking the monthly rent as numeraire

treatment	sample size	Average WTA attribute A	Average WTA attribute B
T3	7	28.5	31.3
T4	8	15.2	3.9

Overall, the behavioral model based on the endowment effect and shifting reference points is in line with the data, whereas the conventional microeconomic model of exogenously given and invariable preferences cannot account for the observed differences. However, the results should be interpreted carefully, as the experiment has several limitations; among these is the lack of salient rewards.

5 Related Work

Literature on negotiation analysis brought up several explanations for empirically observed negotiator behavior. The lack of Pareto-optimal outcomes, for example, is frequently attributed to the following cognitive biases:

- The *fixed pie illusion* states that many negotiators disregard integrative potential and assume to play a constant-sum game [2].
- The *illusion of conflict* implies that parties assume that a compromise which is good for the counterparty is bad for themselves [23].
- The theory of *reactive devaluations* of the counterparty's offers directly follows from the illusion of conflict. Parties devalue any proposal made by the counterparty just because it originates from the counterparty [18].
- Different *fairness perceptions* provide parties to dismiss outcomes which are identified as fair by the counterparty [1].
- The *self-enhancement bias* assumes that negotiators are overly confident in the own skills [11].

The offer-dependent preference construction outlined in the previous sections is not in conflict with these established behavioral patterns and biases but introduces a new perspective on negotiator decision making. All of these effects might interact with each other.

Reference-dependent evaluation of offers—as argued in the present paper—and anchoring are not new to negotiation analysis; anchoring is a common mistake in negotiations [2]. However, studies on anchors in negotiations regard the initial positions of bargainers, i.e. their

opening offers, as anchors. How negotiators adapt anchors and reference points gradually during the process of negotiating is not resolved up to know.

Furthermore, the influence of reference-dependent evaluation on decision making is well-known since Tversky and Kahnemann [24] extended the concept of reference points from risky choices [10] to riskless multi-attribute choices. Tversky and Kahnemann [24] cover the effect of reference-dependence in great detail without addressing the origin of reference points in a multi-attribute decision space. The present paper takes up the established reference point concept and studies its emergence in bilateral negotiations.

Curhan, Neale, and Ross study preferences changing during a negotiation with a focus on dissonance and self-perception theory [6]. The basic idea of endogenously constructed preferences is related to the present study. Yet the authors do not address endowment effects and loss aversion and therefore have a different perspective on a related phenomenon.

Besides the negotiation analytic approaches outlined above, game theory deals with bilateral interactions as well. Game theoretic models of negotiations with reference-dependent preferences are build by Compte and Jehiel [3] and Li [13]; their players' reference points shift as a function of offers received as advocated in Sections 2 and 3. Both papers, however, focus on single-attribute bargaining and the results do not straightforwardly apply to multi-attribute negotiations. Finally, and again for the single-issue case, Kristensen and Gärling find experimental evidence on subjects evaluating offers relative to adaptive reference points [12]. These single-attribute studies strengthen the assumption that related effects might emerge in multi-attribute negotiations as well and preferences might be endogenously constructed during a negotiation process.

6 Conclusion and Outlook

Negotiators oftentimes do not enter a bargaining process with a fixed valuation of alternatives and invariable trade-offs between issues. On the contrary, preferences are endogenously adapted and depend on the history of offers in the ongoing process. The behavioral foundations for this endogenous preference construction—namely reference points, loss aversion, and the endowment effect—were outlined in the paper. Furthermore, data gathered in an internet experiment with students negotiating about tenancy contracts supports the supposition of endogenous preferences.

For negotiation support systems (NSS), the implications are as follows: (1) Systems might warn their users to avoid the sketched bias, (2) they should re-elicit users' preferences at the end of a negotiation if a proposed improvement should be acceptable, and (3) if the changed preference structure is temporary and the (true) ex-ante preferences recur, the system should propose improved agreements with considerable delay after the end of the negotiation. Furthermore, a NSS could assist its user in the offer generation process and might recommend offers counteracting the formation of (virtual) endowment by the counterparty.

Future work will include experiments on negotiations dealing with durable consumer goods and experiments with both parties being played by subjects. These experiments will, contrary to the one presented above, be based on salient rewards. Furthermore, a more sorrow analysis

of implications for negotiation tactics and the design of negotiation support systems is planned.

References

[1] L. Babcock, G. Loewenstein, S. Issacharoff, and C. F. Camerer. Biased judegments of fairness in bargaining. *American Economic Review*, 85(5):1337–1343, 1995.

[2] M. H. Bazerman and M. A. Neale. *Negotiating Rationally*. The Free Press, 1992.

[3] O. Compte and P. Jehiel. Bargaining with reference dependent preferences. Working Paper, 2003.

[4] P. Cramton. Sequential bargaining mechanisms. In A. Roth, editor, *Game-Theoretic Models of Bargaining*, pages 149–179. Cambridge University Press, 1985.

[5] D. Cray and G. Kersten. Negotiating inefficient compromises: Is less better than more? Working Paper, 1999.

[6] J. R. Curhan, M. A. Neale, and L. D. Ross. Dynamic valuation: Preference changes in the context of face-to-face negotiation. *Journal of Experimental Social Psychology*, 40:142–151, 2004.

[7] D. D. Davis and C. A. Holt. *Experimental Economics*. Princeton University Press, 1992.

[8] P. Faratin, C. Sierra, and N. R. Jennings. Using similarity ciretria to make issue trade-offs in automated negotiations. *Artificial Intelligence*, 142:205–237, 2002.

[9] D. Kahneman, J. L. Knetsch, and R. H. Thaler. Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy*, 98(6):1325–1348, 1990.

[10] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

[11] R. M. Kramer, E. Newton, and P. L. Pommerenke. Self-enhancement biases and negotiator judgment: Effects of self-esteem and mood. *Organizational Behavior and Human Decision Processes*, 56(1):110–133, 1993.

[12] H. Kristensen and T. Gärling. Anchor points, reference points, and couteroffers in negotiations. *Group Decision and Negotiation*, 9:453–505, 2000.

[13] D. Li. Bargaining with history-dependent preferences. Working Paper, 2004.

[14] D. G. Pruitt. *Negotiating Behavior*. Academic Press, 1981.

[15] H. Raiffa. *The Art and Science of Negotiation*. Harvard University Press, 1982.

[16] U.-D. Reips. Standards for internet-based experimenting. *Experimental Psychology*, 49(4):243–256, 2002.

[17] U.-D. Reips. Theory and techniques of conducting web experiments. In U.-D. B. M. Batinic, Bernad Reips, editor, *Onlines Social Sciences*, chapter 13, pages 229–250. Hogrefe & Huber Publishers, 2002.

[18] L. Ross and C. Stillinger. Barriers to conflict resolution. *Negotiation Journal*, 7:398–404, 1991.

[19] V. L. Smith. Microeconomic systems as an experimental science. *American Economic Review*, 72(5):923–955, 1982.

[20] M. A. Strahilevitz and G. Loewenstein. The effect of ownership history on the valuation of objects. *Journal of Consumer Research*, 25:276–289, 1998.

[21] M. Tajima and N. M. Fraser. Logrolling procedure for multi-issue negotiation. *Group Decision and Negotiation*, 10:217–235, 2001.

[22] R. H. Thaler. Towards a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, 1(March):39–60, 1980.

[23] L. L. Thompson and R. Hastie. Social perception in negotiation. *Organizational Behavior and Human Decision Processes*, 47:98–123, 1990.

[24] A. Tversky and D. Kahneman. Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, 106:1039–1061, 1991.

[25] R. Vetschera. Preference structures and behavioral consistency in negotiations. Working Paper, 2004.