Theory and Decision manuscript No.
(will be inserted by the editor)

Fairness motivation in bargaining: A matter of principle

Sigbjørn Birkeland · **Bertil Tungodden**

the date of receipt and acceptance should be inserted later

Abstract In this paper, we study the role of fairness motivation in bargaining. We show that bargaining between two strongly fairness motivated individuals who have different views about what represents a fair division may end in disagreement. Further, by applying the Nash bargaining solution, we study the influence of fairness motivation on the bargaining outcome when an agreement is reached. In particular, we show that the bargaining outcome is sensitive to the fairness motivation of the two individuals, unless they both consider an equal division fair. We argue that our results accommodate existing experimental and field data on bargaining.

Keywords Bargaining · Fairness · Disagreement · Nash bargaining solution

Sigbjørn Birkeland
NHH Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway
Tel. +4793480893
E-mail: sigbjorn.birkeland@online.no.

Bertil Tungodden
NHH Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway
Tel. +4755959261
E-mail: Bertil.Tungodden@nhh.no.

1 Introduction

Bargaining is an important mechanism in all realms of society and a prominent topic in the social sciences. Bargaining is typically studied in the context of self-interested individuals, but it is by now well-established that people are also fairness motivated (Camerer, 2003). How does fairness motivation affect the possibility of reaching a bargaining agreement?[1]

Economic experiments using the alternating offer protocol of Rubinstein (1982) have shown that when an agreement is reached, bargaining outcomes tend to cluster around equal division, even when researchers impose unequal bargaining power on the participants (Camerer, 2003). This may suggest that fairness motivation not only affects the possibility of an agreement, but also potentially the nature of an agreement. For example, in a two-person bargaining situation where none of the participants has a particular fairness claim to the endowment, which has been the predominant case in many bargaining experiments, the participants may find it fair to agree on an equal division, even when they have greater bargaining power than the other individual.

The present paper studies formally the importance of fairness motivation in bargaining, and thereby aims at contributing to better understanding of cooperation and conflict in society. We introduce a bargaining model that allows for both self-interest and fairness motivation, and analyze how these two motivational forces interact in the bargaining process. To introduce the possibility of a fairness motivated conflict, we enrich the context such that the individuals do not necessarily find it fair to share equally.[2] Recent economic experiments have, for example, shown that if the endowment to be shared is the result of individual contributions, people may have different views of what constitutes a fair division (Konow, 1996; Gächter and Riedl, 2005; Cappelen, Sørensen, and Tungodden, 2010). In the formal analysis, we provide a framework for studying bargaining situations both when people have compatible and

[1] Political scientists (Hirschman, 1977; Elster, 1989) have nicely captured the potential conflict that may arise between fairness motivated individuals:

> The last case, norm conflict, is less likely to yield negotiated solutions. In norm-free bargaining, the only thing at stake is self-interest, a mild if mean-spirited passion. In norm conflict, the parties argue in terms of their honour, a notoriously strong passion capable of inspiring self-destructive and self-sacrificial behaviour. . . . Compromises are possible between opposing norms, if one or both parties pour some water in their wine and let self-interest override honour. (Elster, 1989, p. 244).

Bargaining disagreement is commonly observed both in the field and in lab experiments, see among others Malouf and Roth (1981); Roth (1995); Ashenfelter and Currie (1990).

[2] Which fairness principle to apply in a particular context is an important and frequently discussed topic in the bargaining literature (Young, 1991). In the experimental literature, it is also commonly acknowledged that a richer fairness context than equal division should be considered. It is, however, sometimes left out for reasons of intractability, as illustrated by the discussion of the dictator game in Andreoni and Bernheim (2009): 'If the players are asymmetric with respect to publicly observed inertia of merit, the fairness of an outcome might depend on the extent to which it departs from some other benchmark, such as $x^F = 0.4$. Provided the players agree on x^F, similar results would follow, except that the behavioral norm would correspond to the alternate benchmark. However, if players have different views of x^F, matters are more complex' (p. 1611, footnote 12). The present paper is also related to Shalev (2002), who analyzes bargaining between individuals who experience loss aversion if the outcome deviates from a reference point, and to de Clippel (2007), who analyzes Nash bargaining solution in terms of disagreement point convexity and midpoint dominance.

incompatible views of what is a fair division of the endowment, and we argue that this framework accommodates existing experimental and field data on bargaining.

In the first part of the paper, we study how fairness motivation influences the possibility of reaching an agreement. Proposition 1 shows that bargaining between two individuals who disagree about what is a fair division of the endowment may end in conflict. This result is shown for a general class of utility functions and illustrates the importance of allowing for heterogeneity in fairness views when studying bargaining. In contrast, Proposition 2 establishes that if two individuals agree on what is a fair division, it is always possible to reach an agreement. We illustrate these results by the use of a specific utility function that represents a standard social preference model (Bolton and Ockenfels, 2000; Bruyn and Bolton, 2008; Cappelen, Hole, Sørensen, and Tungodden, 2007), where we show that for realistic parameter values, fairness motivation can lead to disagreement.

In the second part of the paper, we study the nature of the bargaining agreement by using the Nash bargaining solution. Propositions 3 - 5 show that both the weight attached to fairness and the fairness view are crucial in determining the allocation of the endowment in the bargaining agreement. These results demonstrate how fairness considerations may be important in shaping the bargaining agreement, and may provide some justification for the advice about focussing on fairness put forward in much of the prescriptive bargaining literature (Fisher and Ury, 1991). But the analysis also highlights the fragileness of this advice—if each bargainer insists on a bargaining outcome that reflects what he considers fair, the bargaining may end up in disagreement. Finally, Proposition 6 shows that the introduction of fairness motivation does not affect the bargaining outcome if both individuals consider it fair to divide in proportion to their bargaining power.

2 Experimental evidence

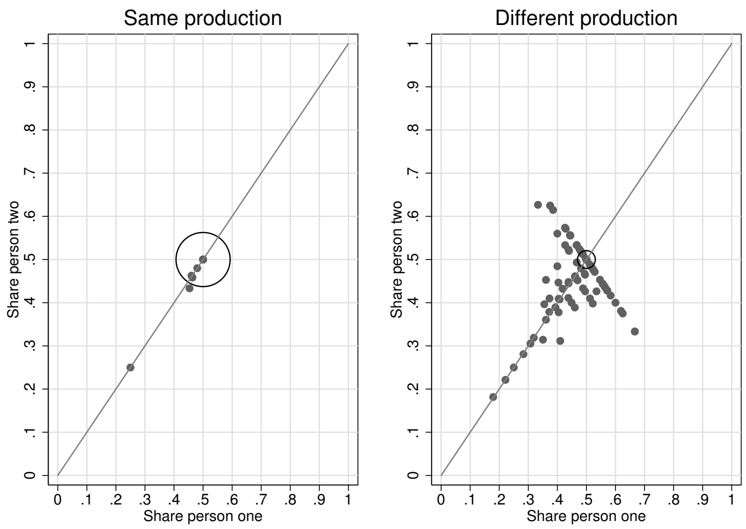
The subgame perfect equilibrium of Rubinstein (1982) predicts an (almost) equal division of the endowment in the non-cooperative infinite alternating offer game where two individuals have a discount rate close to one, even if they are only motivated by self-interest.[3] This is in line with what is justified by the cooperative Nash bargaining solution, which does not rely on a specific bargaining protocol (Nash, 1950).

The prediction of equal division fails dramatically for the alternating offer protocol bargaining experiment reported in Birkeland (2013), where participants differ in their contribution to the endowment. In this experiment, which consists of a production phase and a bargaining phase, the endowment to be shared is a result of individual production. In the production phase, the participants individually produce an output by typing a text from a transcript and receive a monetary reward equal to

[3] The alternating offer game starts with one individual being randomly assigned to be the first mover to propose an opening offer in the first round ($t = 1$). The second mover responds to the opening offer either by accepting it, in which case the bargaining is closed without any costs, or by giving a counter offer in a second round ($t = 2$). The endowment shrinks in each round t by a discount factor δ_i^t. An agreement is reached when one individual accepts the offer from the other individual. The subgame perfect equilibrium of this game uniquely predicts a pair of proposals, which converges to an equal division when there is a common discount factor that approaches one.

each word correctly typed multiplied by a randomly assigned high or low price. In the bargaining phase, the participants are randomly matched into pairs. They are then instructed to bargain over how to spilt the sum of the monetary rewards using the alternating offer bargaining protocol with an infinite horizon, where both participants are induced with the same constant discount factor close to one ($\delta = 0.96$).

Fig. 1 Experimental bargaining results



Note: The left panel shows the bargaining outcomes from 15 situations where the bargainers have the same production and price, and the right panel shows the bargaining outcomes from 97 situations where the bargainers differ in terms of either the amount produced or the price. The circle indicates the number of observations that are an exact 50–50 division in the first round.

Figure 1 presents the bargaining outcomes in this experiment, where the left panel shows the outcomes from the situations where the bargainers have the same production and price and the right panel shows the outcomes from the situations where the bargainers differ in terms of either the amount produced or the assigned price. All the points along the diagonal represent bargaining outcomes where there is an equal division of the endowment.

We observe that when the participants differ either with respect to the amount produced or the assigned price (right panel), the bargaining outcome in the majority of the situations (51%) is an unequal division.[4] Thus, the observed behavior is inconsistent with the prediction of the standard bargaining model in the literature. The most likely explanation is that the bargainers are motivated by fairness considerations that

[4] To accommodate the fact that the participants have to make offers in steps of NOK 5, all bargaining outcomes that give a share in the interval 47.5%-52.5% are characterized as an equal division. The instructions to this experiment are provided in Appendix B.

justify an unequal division, in line with their responses in a post-experimental questionnaire. 96% of the participants found an unequal division justifiable in situations where bargainers differed in terms of either the amount produced or the assigned price, whereas only 4% of the participants found an equal division fair.[5]

In all situations where the participants have identical production and price (left panel), the bargaining outcome is an equal division. These observations are consistent with the standard bargaining solution for individuals that are only motivated by material self-interest, but fairness considerations may still matter in such a context. The fact that everyone agrees that an equal division is fair may make it easier to reach a bargaining agreement, consistent with what we observe in the present experiment.[6] In only one out of 15 situations where the participants have the same production and price does the bargaining process last for more than five rounds, whereas this happens in 22 out of 97 situations when the individuals differ in terms of either the amount produced or the price.

The experimental evidence makes clear that it is important to include fairness motivation in the study of bargaining. In the previous literature, this has been done by introducing social preference models of inequality aversion (Bolton and Ockenfels, 2000), and it has been shown that these models have strong explanatory power in bargaining experiments where none of the participants has a greater fairness claim to the endowment (Bruyn and Bolton, 2008). But models of inequality aversion do not seem to be able to explain bargaining outcomes in cases where people differ in their contributions to the endowment. In such cases, we need to allow for the possibility that people may find it fair to divide unequally. This possibility comes with two fundamental insights for bargaining. First, it becomes likely that people will sometimes disagree with respect to what is the fair division of the endowment, since there are several possible fairness views on how to handle differences in contributions. Second, it may not be feasible for both participants to receive what they perceive a fair share of the endowment, as would be the case if both consider it fair to receive more than half of the endowment. We now turn to a formal analysis of the implications of these insights for the study of bargaining.

3 Theoretical framework

In this section, we introduce a class of utility functions that allow for fairness motivated individuals, and we derive the corresponding set of feasible bargaining agreements. In the following analysis, we do not model a non-cooperative bargaining game, but rather focus on the implications of introducing fairness motivation for the set of feasible agreements (Section 4) and for the Nash bargaining solution (Section 5).

[5] This is also consistent with what is observed in dictator and bargaining games with entitlements, where the outcome typically is an unequal division (Konow, 1996; Cappelen et al, 2010; Gächter and Riedl, 2005).

[6] It may also explain why we see less deviation from equal division than predicted by standard models when inducing unequal discount rates (Ochs and Roth, 1989; Weg, Rapoport, and Felsenthal, 1990).

Consider a bargaining environment in which two individuals, $i = 1,2$, are to divide an endowment, Y, which, without loss of generality, is normalized to $Y = 1$. The individuals can agree on any pair $x = (x_1, x_2)$ of shares of the endowment, $x_i \in [0,1]$, that belongs to the set of feasible bargaining agreements $X = \{x : x_1 + x_2 \leq 1\}$.

An individual is assumed to have preferences that can be represented by a continuous utility function, where the utility derived from an agreement depends on the share of the endowment he receives, how fair he perceives the agreement to be, and the weight he attaches to fairness considerations. Individual i perceives it as fair that he receives the share $s_i \in (0,1]$, and derives disutility from an agreement that he perceives to be unfair. The relative importance attached to fairness considerations is given by $\beta_i \in [0,\infty)$, where $\beta_i = 0$ means that the individual only cares about material self-interest. Thus, $\beta_i > 0$ captures the tension between fairness considerations and material self-interest in an individual's motivation.

In the analysis, we impose the following assumption on the class of utility functions under consideration.

Assumption 1 *Agreement utility. An individual's preferences over the set of feasible agreements are represented by the continuous utility function, $u_i(x_i, s_i, \beta_i)$, which has the following properties:*

(a) $u_i(x_i, s_i, \beta_i)$ is strictly increasing in x_i in the interval $0 \leq x_i \leq s_i$, and
(b) $u_i(x_i, s_i, \beta_i)$ is strictly decreasing in β_i if $|x_i - s_i| > 0$.

In bargaining, an individual evaluates the utility from possible agreements against the utility from a situation where no agreement is reached, which is given by $u_i(d, s_i, \beta_i)$, where d is the disagreement outcome. We make the following assumption about disagreement utility:

Assumption 2 *Disagreement utility. An individual strictly prefers what he perceives as a fair agreement to disagreement, i.e., $u_i(s_i, s_i, \beta_i) > u_i(d, s_i, \beta_i)$, and weakly prefers disagreement to an agreement where he receives nothing, $u_i(d, s_i, \beta_i) \geq u_i(0, s_i, \beta_i)$.*

Define a reservation point, x_i^R, as a share of the endowment that makes the individual indifferent between agreement and disagreement, i.e., $u_i(x_i^R, s_i, \beta_i) = u_i(d, s_i, \beta_i)$. By combining Assumption 1(a) and Assumption 2, it follows that there always exists a reservation point for a continuous utility function, but, as will be shown, there may be more than one reservation point. Let x_i^L denote the lowest reservation point, which means that $u_i(x_i, s_i, \beta_i) < u_i(d, s_i, \beta_i)$ for all $x_i < x_i^L$. We assume that an individual's utility function has the following limit property with respect to the lowest reservation point.

Assumption 3 *Reservation point. The lowest reservation point is strictly and continuously increasing in β_i and approaches in the limit what individual i considers his fair share, i.e., $\lim_{\beta_i \to \infty} x_i^L = s_i$.*

Assumption 3 captures the fact that if an individual attaches more weight to fairness, then he considers more unfair agreements worse than disagreement. In the limit, if he cares only about fairness, then he considers any unfair agreement worse than disagreement.

A bargaining agreement gives the utility pair $(u_1(x_1,s_1,\beta_1),u_2(x_2,s_2,\beta_2))$, where the feasible agreements are given by the set:

$$\mathscr{U} = \{(u_1(x_1,s_1,\beta_1),u_2(x_2,s_2,\beta_2)) : x \in X\}.$$

The bargaining set is defined in the standard way, where a bargaining agreement gives both individuals at least as much utility as no agreement.

Definition 1 Individual rationality. An agreement is in the bargaining set:

$$\mathscr{B}(\mathscr{U}) = \{(u_1(x_1,s_1,\beta_1),u_2(x_2,s_2,\beta_2)) \geq (u_1(d,s_1,\beta_1),u_2(d,s_2,\beta_2)) : x \in X\}.$$

In the analysis, we focus on how the lowest reservation point and the bargaining set are affected by changes in (β_1,β_2), where we then for short refer to them as a functions of these parameter values, $x_i^L(\beta_i)$ and $\mathscr{B}(\beta_1,\beta_2)$.

We allow for the possibility that the individuals may disagree about what they perceive to be the fair allocation of the endowment.[7] Thus, two classes of situations are possible: (i) the individuals follow fairness views that are compatible, where $s_1+s_2=1$ or $s_1+s_2<1$, and (ii) the individuals follow fairness views that are incompatible, $s_1+s_2>1$. In the following analysis, we focus on (i) when $s_1+s_2=1$ and (ii), which represent the more interesting cases. When $s_1+s_2=1$, material self-interest constitutes the source of the bargaining conflict, whereas disagreement about what constitutes the fair share to each of the individuals deepens the bargaining conflict when $s_1+s_2>1$. If $s_1+s_2<1$, then we have a case where both individuals find it fair that the other individual receives more than he himself considers to be his—fair share. We believe that this case is relatively rare in real life where people (at least weakly) seem to have self-serving judgments of fairness (Babcock, Loewenstein, Issacharoff, and Camerer, 1995; Babcock, Wang, and Loewenstein, 1996), so we do not pursue a further discussion of it.

4 Principled agreement and disagreement

We here consider how fairness motivation affects the possibility for reaching an agreement in bargaining. The first part of this section offers some general results, whereas the second part illustrates these results by introducing a specific utility function. The third part provides an illustration of how perceived fair shares can be based on concrete fairness principles in a production context.

4.1 Some general results

Intuitively, one might think that it would be easier for fairness motivated individuals to reach an agreement when bargaining. As shown in Proposition 1, however, this is not necessarily the case. If two individuals hold incompatible fairness views, an agreement may not be feasible.

[7] This may reflect fundamental or contextual differences between the individuals. Akerlof and Kranton (2010) suggest that fairness perceptions may depend on an individual‘s identity, for example, an individual may perceive fairness differently as an employer than as an employee.

Proposition 1 Principled disagreement. *If two individuals hold fairness views that are incompatible, $s_1 + s_2 > 1$, then there exist $(\hat{\beta}_1, \hat{\beta}_2)$ such that for any $(\beta_1 \geq \hat{\beta}_1, \beta_2 \geq \hat{\beta}_2)$, the only feasible solution is the disagreement outcome.*

Proof Principled disagreement.

(1) Consider any fairness views such that $s_1 + s_2 > 1$. It follows that for any $x \in X$, $x_1 \leq s_1 - \varepsilon/2$ or $x_2 \leq s_2 - \varepsilon/2$ or both, where $\varepsilon/2 = s_1 + s_2 - 1$. Without loss of generality, assume that the inequality holds for individual 1 (and denote $\beta_2 = \hat{\beta}_2$).
(2) By the limit property of the lowest reservation point and the fact that it is continuously and strictly increasing in β_1, it follows that there exists $\hat{\beta}_1$ such that $x_1^L(\hat{\beta}_1) > s_1 - \varepsilon/2$.
(3) By (2) and the definition of the lowest reservation point, it follows that $u_1(x_1, s_1, \hat{\beta}_1) < u_1(d, s_1, \hat{\beta}_1)$ for all $x_1 < x_1^L(\hat{\beta}_1)$. Thus $\mathscr{B}(\hat{\beta}_1, \hat{\beta}_2)$ is empty, since an allocation is in the bargaining set only if it gives both individuals at least as much utility as they would get from disagreement. The result now follows from the fact that the lowest reservation point is strictly increasing in β_1.

The result shows, for a general class of utility functions, that principled disagreement may be the result of a bargaining process between fairness minded individuals. In contrast to individuals who are only motivated by material self-interest and who would prefer any share of the endowment to disagreement, a fairness minded individual may prefer disagreement to an unfair share of the endowment. As a result, if two individuals involved in bargaining hold fairness views that are incompatible, then it may not be feasible to find an allocation of the endowment that both find sufficiently fair.

For bargainers who hold perceived fair shares that are compatible, however, it is always possible to reach an agreement.

Proposition 2 Principled agreement. *If two individuals hold fairness views that are compatible, $s_1 + s_2 = 1$, then there always exists a non-empty bargaining set, $\mathscr{B}(\beta_1, \beta_2)$. The bargaining set converges to the allocation that both individuals consider fair when the weight attached to fairness for at least one of the individuals approaches infinity, $\lim_{\beta_i \to \infty} \mathscr{B}(\beta_1, \beta_2) = \{(s_1, s_2)\}$, $i = 1, 2$. The bargaining set is always shrinking with an increase in (β_1, β_2) if the disagreement utility is independent of the weight attached to fairness, $u_i(d, s_i, \beta_i) = u_i(d, s_i, \beta_i^*)$ for all β_i, β_i^*.*

Proof Principled agreement.

(1) Consider any (β_1, β_2) and any fairness views such that $s_1 + s_2 = 1$. It follows that $(s_1, s_2) \in X$. By the fact that an individual strictly prefers what he perceives as a fair agreement to disagreement (Assumption 2), it follows that $(u_1(s_1, s_1, \beta_1), u_2((s_2, s_2, \beta_2))$ $\in \mathscr{B}(\beta_1, \beta_2)$. Thus, the bargaining set is always non-empty.
(2) For any allocation $x \in X \setminus (s_1, s_2)$, it follows that $x_1 < s_1$ or $x_2 < s_2$ or both. From the limit property of the lowest reservation point (Assumption 3), it follows that when β_i goes to infinity, an agreement giving one of the bargainers strictly less than what he perceives as a fair share is not in the bargaining set. By (1), the limit property of the bargaining set follows.

(3) By the properties of the lowest reservation point (Assumption 3), it follows that $(u_1(x_1^L(\beta_1),s_1,\beta_1),u_2(x_2^L(\beta_2),s_2,\beta_2)) \in \mathscr{B}(\beta_1,\beta_2)$. Consider now any $\beta_1^* > \beta_1$ and $\beta_2^* > \beta_2$. By the fact that the lowest reservation point is strictly increasing in β_i (Assumption 3), it follows that $u_i(x_i^L(\beta_i),s_i,\beta_i^*) < u_1(d,s_1,\beta_1^*)$. Thus, $(u_1(x_1^L(\beta_1),s_1,\beta_1^*),u_2(x_2^L(\beta_2),s_2,\beta_2^*)) \notin \mathscr{B}(\beta_1^*,\beta_2^*)$. Further, consider any allocation x that is not in $\mathscr{B}(\beta_1,\beta_2)$. By (1), it follows that for at least one of the individuals, $u_i(x_i,s_i,\beta_i) < u_i(d,s_i,\beta_i)$, where $x_i \neq s_i$. By Assumption 1(b), it follows that $u_i(x_i,s_i,\beta_i) > u_i(x_i,s_i,\beta_i^*)$. By assumption, $u_i(d,s_i,\beta_i) = u_i(d,s_i,\beta_i^*)$, and thus $u_i(x_i,s_i,\beta_i^*) < u_i(x_i,s_i,\beta_i) < u_i(d,s_i,\beta_i) = u_i(d,s_i,\beta_i^*)$. Hence, the allocation x is not in $\mathscr{B}(\beta_1^*,\beta_2^*)$. In sum, the bargaining set shrinks with an increase in (β_1,β_2), since this change removes some allocations from the bargaining set but does not add any new allocations to the bargaining set.

The two propositions show that the bargaining set for fairness minded individuals is typically smaller than the bargaining set for individuals who are only motivated by self-interest. Since a fairness minded individual has disutility from an unfair allocation, the reservation point increases when β_i increases and, consequently, the bargaining set shrinks. In the limit, if the fairness views are compatible, the bargaining set reduces to the only mutually acceptable agreement, the fair division. In contrast, if the fairness views are not compatible, the bargaining set becomes empty if the bargainers places sufficient weight on fairness considerations.

4.2 A specific utility function

We illustrate the analysis by the use of the following standard social preference model for evaluating bargaining agreements,[8]

$$u_i(x_i,s_i,\beta_i) = x_i - \beta_i(x_i - s_i)^2.$$

The utility function attains its inner maximum when:

$$x_i^* = \frac{1}{2\beta_i} + s_i.$$

As $\beta_i \to \infty$, the interior solution approaches the fair share, $x_i^* \to s_i$. On the other hand, if β_i is sufficiently small, the utility function reaches its constrained maximum when the allocation to individual i is equal to the endowment $Y = 1$.

It follows straightforwardly that this utility function satisfies Assumption 1. First, if $0 \leq x_i \leq s_i$, then both terms in the utility function are increasing in x_i, which reflects that the individual derives utility from more income (the first term) and from an allocation coming closer to what he considers a fair division of the endowment (the second term).[9] Second, it follows directly from the second term that the utility function is strictly decreasing in the weight attached to fairness, β_i.

[8] The model is a version of Cappelen et al (2007), see also Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Lopomo and Ok (2001), and Bruyn and Bolton (2008).

[9] The utility function also implies that the individuals are risk neutral, which is commonly assumed in experimental studies where small stakes are involved.

To establish a reservation point, we need to make an assumption about the utility derived from disagreement. In the following, we make the standard assumption that $u_i(d, s_i, \beta_i) = 0$ for all s_i and β_i, which captures the idea that if the bargaining breaks down, both individuals are forced to continue on their own as if the bargaining situation had not happened (Binmore, 2007). Thus, even if fairness plays a role in determining the attractiveness of a potential agreement, we assume that it is not explicitly linked to the disagreement point. This assumption is certainly not uncontroversial, and there might be interesting cases where the disagreement utility is lower when an individual considers his fair share to be large or attaches great importance to fairness considerations.[10]

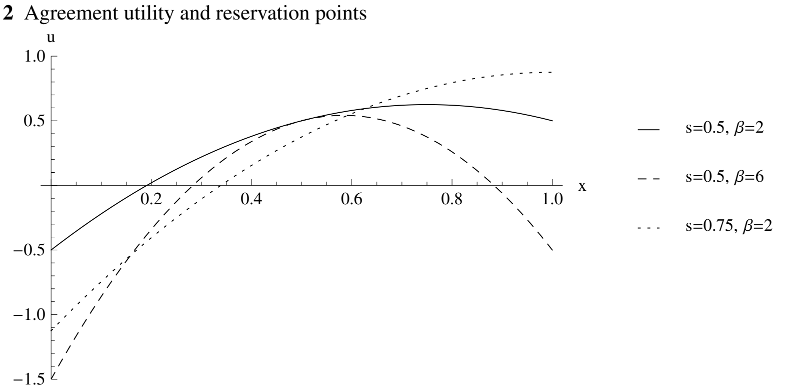
The present formulation of disagreement utility satisfies Assumption 2. An individual strictly prefers what he perceives as a fair agreement to disagreement, since $u_i(s_i, s_i, \beta_i) = s_i > 0$, and weakly prefers disagreement to an agreement where he receives nothing, since $u_i(0, s_i, \beta_i) \leq 0$.

The lower reservation point is given by:

$$x_i^L = \frac{1 + 2\beta_i s_i - \sqrt{1 + 4\beta_i s_i}}{2\beta_i}.$$

The lower reservation point is influenced both by what the individual considers fair and the weight attached to fairness considerations, and it is strictly positive for all individuals with $\beta_i > 0$. In contrast, the lower reservation point for an individual who is only motivated by material self-interest, $\beta_i = 0$, is always zero. Consistent with Assumption 2, the lower reservation point is strictly increasing in β_i and has the limit property that it approaches what the individual perceives to be his fair share.

Fig. 2 Agreement utility and reservation points



Note: The figure shows how the agreement utility $u_i(x_i, s_i, \beta_i)$ and the reservation points relate to different parameter values for the fair share s_i and the weight attached to fairness β_i.

[10] Note that the general analysis in the previous section did not impose this specific assumption, but allowed for the disagreement utility to depend on the fairness view and the weight attached to fairness (except in the last part of Proposition 2, where we assumed that the disagreement utility was independent of the weight attached to fairness).

Figure 2 illustrates how the fairness motive affects the agreement utility and the reservation points. We observe that the utility function attains a negative value for a sufficiently small and sufficiently large x_i, if the weight attached to fairness is sufficiently large. The case with a higher β_i also envelops the case with a lower β_i, reflecting that an increase in fairness motivation only increases the utility loss from deviating from the fair share. Given our assumption about disagreement utility being equal to zero, it follows that the present model allows for more than one reservation point. For large values of β_i, there is an upper reservation point, x_i^H, in the interval $s_i \leq x_i^H \leq 1$. The upper reservation point corresponds to situations where an individual is offered a larger share of the endowment than he considers fair and therefore prefers disagreement, see also Bolton and Ockenfels (2000).[11] Figure 2 also illustrates that the lower reservation point increases by an increase in the weight attached to fairness and by an increase in what the individual perceives as a fair share.

We can now illustrate possible bargaining sets in Figure 3, where the upper panel shows a case where the fairness views are incompatible, and the lower panel shows a case where the individuals hold compatible fairness view. Each line represents the frontier of the bargaining set for specific parameter values of the utility functions. An end-point of the frontier is defined where an individual has maximum utility given that the other individual has enough utility to accept the agreement, which must be at his lower reservation point.

In the lower panel, we observe that the reservation point of an individual may drop below the other individual's utility maximizing offer, and still give the first individual more utility than in the case of disagreement. This occurs when the two bargainers have the same view of fairness, and at least one of them is strongly fairness motivated. The Pareto frontier of the bargaining set is in this case given by the line connecting the two points marked x and y.[12] Pareto optimality is a requirement for the Nash bargaining solution discussed in Section 5.

4.3 Agreement and disagreement: An illustration

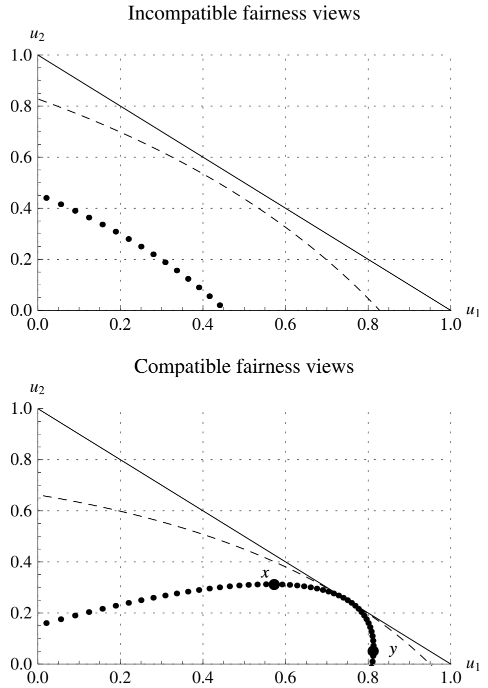
Given that the specific utility function satisfies the assumptions imposed on the general framework, it follows from Proposition 1 that principled disagreement may occur if individuals hold different fairness views. Similarly, it follows from Proposition 2 that if individuals hold compatible fairness views, there will always be a non-empty bargaining set that includes what both consider the fair agreement. We now provide illustrations of these results in an economic environment similar to the one in the experiment discussed in Section 2.

Consider a context where the total endowment is the sum of individual production values, $y_i = e_i p_i$, where $i = 1,2$ and e_i and p_i are the produced amount by individual i and the price assigned to individual i, respectively. Let $e_1 = 1$, $e_2 = \frac{1}{4}$, $p_1 = \frac{1}{4}$, and

[11] In Appendix A, we show that for any fair share, s_i, there is a value $\hat{\beta}_i$ such that for any $0 < \beta_i < \hat{\beta}_i$ there exists a unique reservation point, x_i^L, and for any $\beta_i \geq \hat{\beta}_i$ there exist two reservation points, x_i^L, and x_i^H.

[12] The point marked x in Figure 3 is defined by the maximum utility that individual 2 can achieve, $(\tilde{u}_1, u_2^{max})$, where $\tilde{u}_1 > u_1(x^L) = 0$, and the point marked y is defined correspondingly.

Fig. 3 Bargaining sets



Note: The upper panel shows the frontier of bargaining sets where individuals hold incompatible fairness views ($s_1 = \frac{3}{4}$, $s_2 = \frac{3}{4}$), and the lower panel shows the frontier of bargaining sets where they hold compatible fairness views ($s_1 = \frac{3}{4}$, $s_2 = \frac{1}{4}$). The lines represent different values of β_i (solid line: $\beta_1 = \beta_2 = 0$, dashed line: $\beta_1 = \beta_2 = 0.5$, dotted line $\beta_1 = \beta_2 = 4$). In the lower panel, x and y are the end-points of the Pareto frontier for the dotted line.

$p_2 = 3$, which gives the production values $y_1 = \frac{1}{4}$ and $y_2 = \frac{3}{4}$. This economic environment may, for example, illustrate a situation where two executives bargain about their share of a bonus in a corporation where one business area is exposed to the oil price and another business area is exposed to the aluminium price. In this context, three fairness principles are salient, $k = E, L, P$. First, *strict equality*, which implies a fair share $s_{i(E)} = \frac{1}{2}$. Second, a *laissez-faire* principle, where the individual production values determine the fair shares, $s_{i(L)} = y_i$. Third, a *proportionality* principle, where the fair share is proportional to production, $s_{i(P)} = \frac{e_i}{e_1+e_2}$, which implies that prices are considered irrelevant in a fairness consideration. In this example, if the two individuals hold a fairness view that reflects one of these fairness principles, nine different combinations of fair shares may influence the bargaining process, as illustrated in Table 1.

To illustrate the possibility for principled disagreement, consider the case from Table 1 where $s_{1(P)} = \frac{4}{5}$ and $s_{2(L)} = \frac{3}{4}$, which implies that they are incompatible. If both individuals have $\beta_2 = 8$, their lower reservation points are $x_1^L = 0.54$ and $x_2^L = 0.5$. Consequently, the bargaining set is empty. Interestingly, recent studies suggest

Table 1 Combinations of fairness views: An illustration

	$s_{2(E)}$	$s_{2(L)}$	$s_{2(P)}$
$s_{1(E)}$	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, \frac{3}{4})$	$(\frac{1}{2}, \frac{1}{5})$
$s_{1(L)}$	$(\frac{1}{4}, \frac{1}{2})$	$(\frac{1}{4}, \frac{3}{4})$	$(\frac{1}{4}, \frac{1}{5})$
$s_{1(P)}$	$(\frac{4}{5}, \frac{1}{2})$	$(\frac{4}{5}, \frac{3}{4})$	$(\frac{4}{5}, \frac{1}{5})$

Note: The table shows combinations of fairness views for individuals 1 and 2. The fairness views are given by the fairness principles strict equality (E), laissez-faire (L), and proportionality (P), where individual production value is given by $y_i = e_i p_i$ and $e_1 = 1, e_2 = \frac{1}{4}, p_1 = \frac{1}{4}$, and $p_2 = 3$. Each entry shows what individual 1 and individual 2 perceive as a fair share for themselves.

that $\beta_2 = 8$ is not an extreme case. Bruyn and Bolton (2008) and Cappelen et al (2007) estimate the average weight that an individual attaches to fairness in a three-round bargaining game and a dictator game to be 6.0 and 7.7, respectively.[13] This suggests that for reasonable parameter values, the model predicts disagreement in a bargaining process if the individuals have fairness views that are incompatible.

In line with Proposition 2, if two individuals follow the same fairness principle, there always exists a non-empty bargaining set. For example, consider the case from Table 1 where both individuals consider it fair that individual 1 receives one-fourth of the endowment, i.e., $s_{1(L)} = \frac{1}{4}$ and $s_{2(L)} = \frac{3}{4}$. In this case, if $\beta_1 = \beta_2 = 8$, it follows that the bargaining set contains all feasible divisions of the endowment where $0.1 < x_1 < 0.5$ and $0.5 < x_2 < 0.9$, including what both individuals consider the fair agreement.

5 The nature of the bargaining agreement

In this section, we study how fairness motivation affects the nature of a bargaining agreement when individuals can be represented by the continuous and differentiable utility function introduced in Section 4.

5.1 The Nash bargaining solution

The Nash bargaining solution is commonly used in the analysis of two-person bargaining, and it can be given both a cooperative and a non-cooperative justification. The Nash bargaining solution (x_1^N, x_2^N) is found by maximizing the product of the gain from bargaining for the two individuals, where the gain is given by the difference between the agreement utility and the disagreement utility.[14]

[13] We have here converted their estimates to make them comparable in the present framework.

[14] Nash (1950) characterizes this solution by the following four axioms: (i) independence of affine transformations of the utility function, (ii) independence of irrelevant alternatives, (iii) symmetric treatment of individuals, and (iv) Pareto optimality. Rubinstein (1982) provides a non-cooperative justification for the Nash bargaining solution, see also Binmore, Rubinstein, and Wolinsky (1986). For a further discussion of Nash bargaining theory, see Roth (1979) and Binmore (2007).

$$\max_{x_1} \quad (u_1(x_1,s_1,\beta_1) - u_1(d,s_1,\beta_1))(u_2(x_2,s_2,\beta_2) - u_2(d,s_2,\beta_2))$$
$$\text{s.t.} \quad x_1 + x_2 = 1.$$

We assume that $u_i(d,s_i,\beta_i) = 0$, and thus the first order condition of the maximization problem is given by:

$$\frac{u_1'(x_1,\cdot)}{u_2'(1-x_1,\cdot)} = \frac{u_1(x_1,\cdot)}{u_2(1-x_1,\cdot)}.$$

The analytical solution to the Nash bargaining problem with the specific utility function introduced in Section 4.2 is provided in the appendix. In the following, we study how the weight attached to fairness and the perceived fair share affect the first order condition, and thereby the Nash bargaining solution, in an interesting set of situations. In this analysis, we assume that the bargaining set is non-empty. We also consider implications of allowing for asymmetric bargaining power, by introducing the generalized Nash bargaining solution.

5.2 Fairness motivation and the Nash bargaining solution

We first consider the Nash bargaining solution for a set of cases where one of the individuals attaches more weight to fairness and has a larger perceived fair share than the other.

Proposition 3 Incompatible fairness views and the Nash bargaining solution. *Consider two individuals with the utility function $u_i(x_i,s_i,\beta_i) = x_i - \beta_i(x_i - s_i)^2$ and disagreement utility $u_i(d,s_i,\beta_i) = 0$, $i = 1,2$. If (i) $\beta_1 \geq \beta_2 > 0$, (ii) $s_1 \geq s_2 > 1/2$, and (iii) the comparison between 1 and 2 in (i) or (ii) holds with strict inequality, then the Nash bargaining solution assigns a larger share of the endowment to individual 1, $x_1^N > x_2^N$.*

Proof Incompatible fairness views and the Nash bargaining solution.

(1) By assumption, (i) $\beta_1 \geq \beta_2 > 0$, (ii) $s_1 \geq s_2 > 1/2$, and (iii) the comparison between 1 and 2 in (i) or (ii) holds with strict inequality.
(2) Suppose that $x_2^N \geq x_1^N$ and $x_2^N > s_2$. It then follows that $x_1^N < s_1$ and that $u_1'(x_1^N,\cdot) > u_2'(1-x_1^N,\cdot)$. (Here and later we evaluate the derivatives of the specific utility function introduced in Section 4.2). Hence, it follows from the first order condition of the Nash bargaining solution that $u_1(x_1^N,\cdot) > u_2(1-x_1^N,\cdot)$. But consider now the allocation $x_2 = s_2$ and $x_1 = (1-s_2)$. It follows that $u_1(1-s_2,\cdot)u_2(s_2,\cdot) > u_1(x_1^N,\cdot)u_2(1-x_1^N,\cdot)$, since $u_1(1-s_2,\cdot) > u_1(x_1^N,\cdot)$ and $u_2(s_2,\cdot) > u_1(x_1^N,\cdot) > u_2(1-x_1^N,\cdot)$. Thus, (x_1^N, x_2^N) does not maximize the Nash bargaining product if $x_2^N > s_2$, and the supposition in the first sentence of this part of the proof is false.
(3) Suppose that $x_2^N \geq x_1^N$ and $x_2^N \leq s_2$. This implies that $u_1(x_1^N,\cdot) < u_2(1-x_1^N,\cdot)$ and $u_1'(x_1^N,\cdot) > u_2'(1-x_1^N,\cdot)$, which is inconsistent with the first order condition of the Nash bargaining solution. The supposition in the first sentence of this part of the proof is therefore false and the result follows from combining (2) and (3).

Proposition 3 shows that if the two individuals have incompatible fairness views and one of the individuals dominates the other in both fairness dimensions, then the Nash bargaining solution assigns a larger share of the endowment to this individual. This suggests that to be strongly attached to a fairness view that justifies a large share to oneself strengthens one's bargaining position. If, on the other hand, one individual attaches greater weight to fairness and the other individual has a greater perceived fair share, then it is not in general clear who receives a larger share of the endowment in the Nash bargaining solution. However, if we restrict ourselves to cases where the two individuals have compatible fairness views, then we can establish the following proposition.

Proposition 4 Compatible fairness views and the Nash bargaining solution. *Consider two individuals with the utility function $u_i(x_i, s_i, \beta_i) = x_i - \beta_i(x_i - s_i)^2$ and disagreement utility $u_i(d, s_i, \beta_i) = 0$, $i = 1, 2$. If $s_1 + s_2 = 1$, then the Nash bargaining solution is in the interval between a perceived fair division (s_1, s_2) and and equal division $(1/2, 1/2)$.*

Proof Compatible fairness views and the Nash bargaining solution.

(1) By assumption, $s_1 + s_2 = 1$. Consider the situation where $s_1 = s_2 = 1/2$ and the allocation $x_1^N = x_2^N = 1/2$. It follows that $u_1(x_1^N, \cdot) = u_2(1 - x_1^N, \cdot)$ and $u_1'(x_1^N, \cdot) > u_2'(1 - x_1^N, \cdot)$, and thus the first order condition is satisfied. By uniqueness of the generalised Nash bargaining solution, the result follows.
(2) Consider $s_1 > s_2$. By exactly the same argument as (2) in the proof of Proposition 3, it follows that $x_1^N \leq s_1$ and $x_2^N \geq s_2$. Suppose that $x_1^N < 1/2$. It then follows that $u_1'(x_1^N, \cdot) > u_2'(1 - x_1^N, \cdot)$. Hence, it follows from the first order condition of the Nash bargaining solution that $u_1(x_1^N, \cdot) > u_2(1 - x_1^N, \cdot)$. But consider now the allocation $x_1 = x_2 = 1/2$. It follows that $u_1(1/2, \cdot)u_2(1/2, \cdot) > u_1(x_1^N, \cdot)u_2(1 - x_1^N, \cdot)$, since $u_1(1/2, \cdot) > u_1(x_1^N, \cdot)$ and $u_2(1/2, \cdot) > u_1(x_1^N, \cdot) > u_2(1 - x_1^N, \cdot)$. Thus, (x_1^N, x_2^N) does not maximize the Nash bargaining product if $x_1^N < 1/2$ and the supposition is therefore false. The result follows.

We now turn to a discussion of how an increase in either the weight attached to fairness or the perceived fair share of an individual affects the Nash bargaining solution. As shown by next proposition, both changes improve the bargaining outcome of the individual if he initially receives less than what he considers his fair share.

Proposition 5 Changes in fairness motivation and the Nash bargaining solution. *Consider two individuals with the utility function $u_i(x_i, s_i, \beta_i) = x_i - \beta_i(x_i - s_i)^2$ and disagreement utility $u_i(d, s_i, \beta_i) = 0$, $i = 1, 2$. If $x_1^N < s_1$, then an increase in β_1 or s_1 increases the amount of the endowment assigned to individual 1 in the Nash bargaining solution.*

Proof Changes in fairness motivation and the Nash bargaining solution.

(1) By assumption, $x_1^N(\hat{\beta}_1, \cdot) = \hat{x}_1 < s_1$ when $\beta_1 = \hat{\beta}_1$. Consider now $\beta_1^* > \hat{\beta}_1$. Evaluating the different terms in the first order condition of the Nash bargaining solution, we observe that $u_2(1 - \hat{x}_1, \cdot)$ and $u_2'(1 - \hat{x}_1, \cdot)$ are unaffected by such a

change. For individual 1, however, $u_1(\hat{x}_1, \beta_1^*, \cdot) < u_1(\hat{x}_1, \hat{\beta}_1, \cdot)$ and $u_1'(\hat{x}_1, \beta_1^*, \cdot) > u_1'(\hat{x}_1, \hat{\beta}_1, \cdot)$. Hence, $\frac{u_1(\hat{x}_1, \beta_1^*, \cdot)}{u_2(1-\hat{x}_1, \cdot)} < \frac{u_1'(\hat{x}_1, \beta_1^*, \cdot)}{u_2'(1-\hat{x}_1, \cdot)}$, which implies that the first order condition of the Nash bargaining solution is not satisfied. Hence, $x_1^N(\beta_1^*, \cdot) = x_1^* \neq \hat{x}_1$.

(2) Suppose that $x_1^* < \hat{x}_1$. By the same argument as (2) in the proof of Proposition 3, we can rule out that $(1-\hat{x}_1) < s_2$. Thus, it follows that $\frac{u_1(x_1^*, \beta_1^*, \cdot)}{u_2(1-x_1^*, \cdot)} < \frac{u_1(\hat{x}_1, \beta_1^*, \cdot)}{u_2(1-\hat{x}_1, \cdot)}$ and $\frac{u_1'(x_1^*, \beta_1^*, \cdot)}{u_2'(1-x_1^*, \cdot)} > \frac{u_1'(\hat{x}_1, \beta_1^*, \cdot)}{u_2'(1-\hat{x}_1, \cdot)}$. Taking into account (1), this implies that $\frac{u_1(x_1^*, \beta_1^*, \cdot)}{u_2(1-x_1^*, \cdot)} < \frac{u_1'(x_1^*, \beta_1^*, \cdot)}{u_2'(1-x_1^*, \cdot)}$. But this violates the first order condition of the Nash bargaining solution. The supposition is therefore false and the result follows.

(3) Exactly the same argument as in $(1)-(2)$ applies to an increase in s_1.

We close this section by considering how bargaining power interacts with fairness motivation in the generalized Nash bargaining solution (x_1^{GN}, x_2^{GN}), where bargaining power is captured by each individual's gain from bargaining being raised to the power of α_i. In general, greater bargaining power implies a larger share of the endowment. If people are fairness motivated, however, the presence of bargaining power also makes the weight attached to fairness irrelevant in an interesting set of cases.

Proposition 6 Fairness weight impotency. *Consider two individuals with the utility function* $u_i(x_i, s_i, \beta_i) = x_i - \beta_i(x_i - s_i)^2$ *and disagreement utility* $u_i(d, s_i, \beta_i) = 0$, $i = 1,2$. *If* $s_1 + s_2 = 1$, $\alpha_1 = s_1$, *and* $\alpha_2 = s_2$, *then the generalized Nash bargaining solution is* $(x_1^{GN} = s_1, x_2^{GN} = s_2)$ *for any* β_1, β_2.

Proof Fairness weight impotency.[15]

(1) By assumption, $s_1 + s_2 = 1$, $\alpha_1 = s_1$, and $\alpha_2 = s_2$. The generalized bargaining solution can be found by solving the following optimization problem:

$$\begin{aligned} \max_{x_1} \quad & u_1(x_1, s_1, \beta_1)^{\alpha_1} u_2(x_2, s_2, \beta_2)^{\alpha_2} \\ \text{s.t.} \quad & x_1 + x_2 = 1. \end{aligned}$$

(2) Substitute the constraint into the objective function and perform a logarithmic transformation of the function:

$$\max f(x_1) = \alpha_1 \log u_1(x_1, s_1, \beta_1) + \alpha_2 \log u_2(x_2, s_2, \beta_2).$$

(3) By differentiating f with respect to x_1 and substituting s_1 for α_1 and $1 - s_1$ for α_2, we can write the first order condition as follows,

$$\frac{s_1 - 2\beta_1 s_1 (x_1 - s_1)}{x_1 - \beta_1 (x_1 - s_1)^2} + \frac{(1 - s_1)(-1 + 2\beta_2 (s_1 - x_1))}{1 - x_1 - \beta_2 (s_1 - x_1)^2} = 0,$$

which is satisfied for $x_1 = s_1$, and $x_2 = s_2$. By uniqueness of the generalised Nash bargaining solution, the result follows.

[15] We would like to thank an anonymous referee for suggesting the structure of this proof.

Proposition 6 shows that allocation of bargaining power in proportion to the perceived fair shares can secure a bargaining outcome that is in line with the fairness view of the individuals. Clearly, this is not a feasible approach when individuals disagree about what is a fair division. In such cases, as follows from Proposition 1, no allocation of bargaining power can ensure that the bargaining does not end in principled disagreement.

An immediate corollary of Proposition 6 is that the introduction of fairness motivation does not affect the Nash bargaining solution when both individuals consider an equal division as fair, in line with what was established in the first part of the proof of Proposition 4.

Corollary 1 Irrelevance of fairness motivation. *Consider two individuals with the utility function* $u_i(x_i, s_i, \beta_i) = x_i - \beta_i(x_i - s_i)^2$ *and disagreement utility* $u_i(d, s_i, \beta_i) = 0$, $i = 1, 2$. *If both individuals consider an equal division as fair,* $s_1 = s_2 = \frac{1}{2}$, *then the Nash bargaining solution is* $(x_1^N = \frac{1}{2}, x_2^N = \frac{1}{2})$ *for any* β_1, β_2.

Proof In the Nash bargaining solution, $\alpha_1 = \alpha_2 = 1/2$. By assumption, $s_1 = s_2 = 1/2$, and thus the conditions in Proposition 6 are satisfied. The result follows.

The Nash bargaining solution with self-interested individuals is an equal division of the endowment. The corollary shows that for fairness motivation to make a difference in this model, at least one of the individuals would need to consider it unfair to divide equally. As we have argued earlier, this is commonly the case in complex bargaining situations, and thus we believe that the propositions presented in this section provide insights of importance for many real life situations.

5.3 Numerical examples

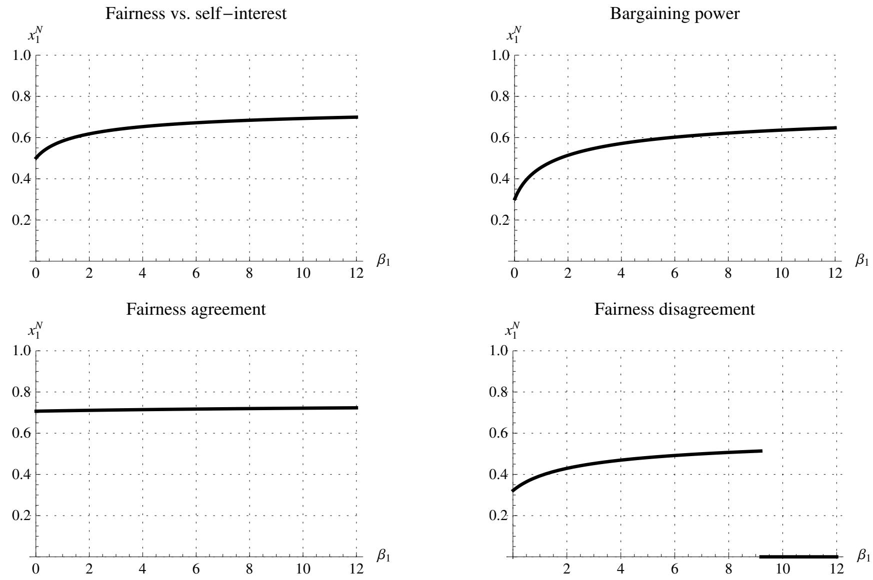
We here illustrate how different combinations of fairness motivation affect the Nash bargaining solution.

Let us first consider the Nash bargaining solution when fairness views are compatible, but justify an unequal division. This is illustrated in the two left panels in Figure 4, where both individuals agree that the fair division is $s_1 = \frac{3}{4}$ and $s_2 = \frac{1}{4}$. We observe that in these cases, the Nash bargaining solution is in the interval between an equal division and what the individuals consider a fair division, in line with Proposition 4. In the upper left panel, we observe that if both individuals are only motivated by self-interest, the outcome is an equal division. When the weight attached to fairness by individual 1 increases, however, the outcome approaches what he perceives to be the fair division, in line with Proposition 2.[16] The convergence to a fair division is strengthened if individual 2 also attaches great weight to fairness, as illustrated in the lower left panel.

Second, let us consider the case where the fairness views are incompatible. In the lower right panel in Figure 4, both individuals consider it fair to receive three-fourths

[16] In Appendix A, we provide the analytical solution for the case where one individual is fairness motivated and the other is only motivated by self-interest.

Fig. 4 The Nash bargaining solution



Note: The panels show what individual 1 receives in the Nash bargaining solution, x_1^N, for different values of β_1. Individual 1 has $s_1 = \frac{3}{4}$ in all the panels. The parameter values for individual 2 are: 'Fairness vs. self-interest' and 'Bargaining power': $\beta_2 = 0$, $s_2 = \frac{1}{4}$ (but the fairness view of individual 2 is irrelevant in these cases), and $\alpha_2 = 0.7$ in 'Bargaining power'; 'Fairness agreement': $\beta_2 = 7$, $s_2 = \frac{1}{4}$; 'Fairness disagreement': $\beta_2 = 7$, $s_2 = \frac{3}{4}$.

of the endowment. We observe that the bargaining outcome assigns a larger share of the endowment to the individual that attaches most weight to fairness, in line with Proposition 4. In line with Proposition 1, however, if individual 1 attaches sufficiently great weight to fairness, the bargaining outcome ends in principled disagreement.

Finally, in the upper right panel, both individuals start out with a low fairness weight, but individual 2 is allocated a larger share because of greater bargaining power. When individual 1 attaches sufficient weight to fairness, however, the Nash bargaining solution assigns a larger share to individual 1. This illustrates that strong fairness motivation may compensate for little bargaining power.

6 Concluding remarks

Standard bargaining models predict that self-interested individuals would reach agreement in bargaining, since they prefer even a small share of the endowment to disagreement. In contrast, we have shown that disagreement may well be the outcome of bargaining between fairness minded individuals. If people hold incompatible fairness views, it will be impossible to reach an agreement if they attach sufficient weight to fairness. People who are motivated by the same fairness view, on the other hand, always face a non-empty bargaining set, and the bargaining outcome approaches the fair division when the bargainers attach great weight to fairness. By using the Nash bargaining solution, we have shown that also the nature of the bargaining agreement is influenced by the introduction of fairness motivation, unless both individuals consider it fair to divide equally.

The formal analysis fits nicely the experimental data presented in Section 2; it predicts an equal division when equality is the only salient fairness view and an unequal division when individuals consider equality unfair. Furthermore, the prediction that the bargaining outcome should be between an equal division and what the individuals consider a fair division is consistent with what is observed in the experiment; in 96 out of 97 situations where the individuals do not have identical production, the bargaining outcome is between an equal division and what would be a fair division according to other salient fairness principles. We did not observe, strictly speaking, principled disagreement in the experiment, but observe that the bargaining, on average, lasted longer when the individuals differed in their production or the assigned price.

This research could be extended both theoretically and empirically. An interesting theoretical extension would be to study in more detail how different assumptions about the relationship between fairness motivation and disagreement utility affect the bargaining outcome. In the second part of the paper, we have focussed on the case where disagreement utility is independent of the weight attached to fairness and what is considered a fair division, but one may certainly think of cases where this assumption does not necessarily apply. Furthermore, it would be interesting to extend the analysis to also consider more formally how fairness motivation affects the efficiency of the bargaining processes. We have shown that greater weight on fairness typically shrinks the bargaining set, but have not provided any theoretical predications of whether this should make it easier or harder to reach an agreement in the bargaining.

An interesting empirical extension would be to study more carefully in the lab how fairness motivation develops in a non-cooperative bargaining process, where the concern for fairness may possibly weaken in later rounds. This may explain the pattern observed in the experimental data, where an equal division (the prediction if individuals only care about material self-interest) is the typical outcome when the bargaining lasts for many rounds. Finally, we would like to point to several empirical hypotheses outside the lab that are suggested by our analysis. First, material self-interest is more predominant in societies where there is substantial heterogeneity in fairness views (since fostering self-interest may then reduce the likelihood of bargaining conflicts), and, conversely, in more homogeneous societies, people are to a greater extent fairness motivated (since fostering fairness motivation may then facilitate reaching bargaining agreements). Second, there are more bargaining conflicts in societies where powerful groups hold different fairness views (since heterogeneity in fairness views increases the likelihood of conflict), for example in societies where employers and employees strongly disagree on what is a fair wage. Third, fairness motivation is stronger in groups with weak bargaining power (since increased concern for fairness may compensate for little bargaining power), for example a labor union in wage negotiations.

Acknowledgements We would like to thank the editor Arno Riedl, three anonymous referees, Alexander W. Cappelen, Kristian Horn, James Konow, Shachar Kariv, and participants at the 5th Nordic Conference on Behavioural and Experimental Economics at Aalto University for valuable comments and helpful discussions. This research project was administered by The Choice Lab.

References

Akerlof GA, Kranton RE (2010) Identity Economics. Princeton University Press

Andreoni J, Bernheim D (2009) Social image and the 50-50 norm: A theoretical and experiemental analysis of audience effects. Econometrica 77:1607–1636

Ashenfelter O, Currie J (1990) Negotiator behavior and the occurrence of disputes. American Economic Review 80(2):416–420

Babcock L, Loewenstein G, Issacharoff S, Camerer C (1995) Biased judgements of fairness in bargaining. American Economic Review 85(5):1337–1343

Babcock L, Wang X, Loewenstein G (1996) Choosing the wrong pond: Social comparisons in negotiations that reflect a self-serving bias. The Quarterly Journal of Economics CXI(1):1–19

Binmore K (2007) Playing for Real. Oxford University Press

Binmore K, Rubinstein A, Wolinsky A (1986) The Nash bargaining solution in economic modelling. Rand Journal of Economics 17:176–188

Birkeland S (2013) Negotiation under possible third party resolution. Journal of Law and Economics 56, forthcoming

Bolton GE, Ockenfels A (2000) Erc: A theory of equity, reciprocity and competition. American Economic Review 90:166–193

Bruyn AD, Bolton GE (2008) Estimating the influence of fairness on bargaining behavior. Management Science 54(10):1774–1791

Camerer C (2003) Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press

Cappelen AW, Hole AD, Sørensen EØ, Tungodden B (2007) The pluralism of fairness ideals: An experimental approach. American Economic Review 97(3):818–827

Cappelen AW, Sørensen EØ, Tungodden B (2010) Responsibility for what? Fairness and individual responsibility. European Economic Review 54(3):429–441

de Clippel G (2007) An axiomatization of the Nash bargaining solution. Social Choice and Welfare 29:201–210

Elster J (1989) The Cement of Society. Cambridge University Press

Fehr E, Schmidt KM (1999) A theory of fairness, competition and cooperation. The Quarterly Journal of Economics 114(3):917–868

Fisher R, Ury W (1991) Getting to YES – Negotiating Agreement Without Giving In, 2nd edn. Penguin Books

Gächter S, Riedl A (2005) Moral property rights in bargaining with infeasible claims. Management Science 51(2):249–263

Hirschman AO (1977) The Passions and the Interests : Political Arguments for Capitalism before Its Triumph. Princeton University Press

Konow J (1996) A positive theory of economic fairness. Journal of Economic Behaviour and Organisation 31:13–35
Lopomo G, Ok EA (2001) Bargaining, interdependence, and the rationality of fair division. RAND Journal of Economics 32:263–283
Malouf MWK, Roth AE (1981) Disagreement in bargaining: An experimental study. Journal of Conflict Resolution 25(2):329–348
Nash JF (1950) The bargaining problem. Econometrica 18:155–62
Ochs J, Roth AE (1989) An experimental study of sequential bargaining. American Economic Review 79(3):355–384
Roth AE (1979) Axiomatic Models of Bargaining. Lecture Notes in Economics and Mathematical Systems, Springer-Verlag
Roth AE (1995) Bargaining experiments. In: The Handbook of Experimental Economics, Princeton University Press, pp 253–348
Rubinstein A (1982) Perfect equilibrium in a bargaining model. Econometrica 50:97–109
Shalev J (2002) Loss aversion and bargaining. Theory and Decision 52:201–232
Weg E, Rapoport A, Felsenthal DS (1990) Two-person bargaining behavior in fixed discounting factors game with infinite horizon. Games and Economic Behavior 2:76–95
Young HP (1991) Fair Division, The University of Michigan Press, chap 2, pp 25–45

Appendix A Analytical results

We here present analytical results referred to in the main text.

Reservation points

As referred to in footnote 10, we here show that for any fair share, s_i, there is a value $\hat{\beta}_i$ such that for any $0 < \beta_i < \hat{\beta}_i$ there exists a unique reservation point, x_i^L, and for any $\beta_i \geq \hat{\beta}_i$ there exist two reservation points, x_i^L, and x_i^H. The utility function,

$$u_i(x_i, s_i, \beta_i) = x_i - \beta_i (x_i - s_i)^2,$$

is a quadratic equation,

$$a_1 x^2 + a_2 x + a_3 = 0,$$

where the coefficients are reduced to:

$$\begin{aligned} a_1 &= -\beta_i, \\ a_2 &= (1 + 2\beta_i s_i), \\ a_3 &= -\beta_i (s_i)^2. \end{aligned}$$

The discriminant, $a_2^2 - 4a_1 a_3 = 1 + 4\beta_i s_i$, is positive and hence the utility function has two real, distinct roots. The quadratic formula gives the two solutions:

$$x_i^L = \frac{1 + 2\beta_i s_i - \sqrt{1 + 4\beta_i s_i}}{2\beta_i}, \qquad x_i^H = \frac{1 + 2\beta_i s_i + \sqrt{1 + 4\beta_i s_i}}{2\beta_i}.$$

These solutions are not defined for $\beta_i = 0$. By definition, a fair share, s_i, can have values in the interval $0 \leq s_i \leq 1$. We observe that if $s_i = 0$, then $x_i^L = 0$. Differentiate x_i^L with respect to β_i:

$$\frac{dx_i^L}{d\beta_i} = \frac{1 + 2\beta_i s_i - \sqrt{1 + 4\beta_i s_i}}{2\beta_i^2 \sqrt{1 + 4\beta_i s_i}}.$$

For the numerator to be positive, $1 + 2\beta_i s_i > \sqrt{1 + 4\beta_i s_i}$. By squaring both sides of the inequality we find that the numerator is always positive for $s_i > 0$. Hence, $\frac{dx_i^L}{d\beta_i} > 0$, and x_i^L is strictly increasing in β_i for $s_i > 0$. We know from Section 3.3 that $\lim_{\beta_i \to \infty} x_i^L = s_i$. Thus, there always exists a lower reservation point, x_i^L, which attains values in the interval $0 \leq x_i^L \leq s_i$.

We then consider the upper reservation point, x_i^H. We observe that if $s_i = 1$, then $x_i^H > 1$, which is outside of the domain of the utility function for argument x_i. We now differentiate x_i^H with respect to β_i:

$$\frac{dx_i^H}{d\beta_i} = \frac{-1 - 2\beta_i s_i - \sqrt{1 + 4\beta_i s_i}}{2\beta_i^2 \sqrt{1 + 4\beta_i s_i}}.$$

We observe that x_i^H is strictly decreasing in β_i, since $\frac{dx_i^H}{d\beta_i} < 0$. If $s_i < 1$, then x_i^H may attain values in the interval $0 < x_i^H \leq 1$, depending on the relationship between s_i and β_i. Define $\hat{\beta}_i$ such that $x^H = 1$, which gives:

$$\hat{\beta}_i = \frac{1}{(1 - s_i)^2}.$$

Any $\beta_i \geq \hat{\beta}_i$ will give an upper reservation point in the interval $0 < x_i^H \leq 1$. Hence, for β_i in the interval $0 < \beta_i < \hat{\beta}_i$, there exists a unique reservation point, x_i^L, and for $\beta_i \geq \hat{\beta}_i$ there exist two reservation points, x_i^L, and x_i^H.

The Nash bargaining solution

As refereed to in Section 5, we here provide the analytical solution to the Nash bargaining problem. The Nash bargaining solution can be found by solving the optimization problem:

$$\begin{aligned} \max_{x_1} \quad & (u_1(x_1, s_1, \beta_1) - u_1(d, s_1, \beta_1))(u_2(x_2, s_2, \beta_2) - u_2(d, s_2, \beta_2)) \\ \text{s.t.} \quad & x_1 + x_2 = 1. \end{aligned}$$

We assume that $u_i(d, s_i, \beta_i) = 0$ and consider the utility function $u_i(x_i, s_i, \beta_i) = x_i - \beta_i (x_i - s_i)^2$. By substituting the constraint into the objective function, the optimization problem can be written as:

$$\max f(x_1) = \left(x_1 - \beta_1 (x_1 - s_1)^2\right)\left(1 - x_1 - \beta_2 (1 - x_1 - s_2)^2\right).$$

Differentiating f with respect to x_1 gives a cubic equation (the subscript on x is suppressed):

$$a_1 x^3 + a_2 x^2 + a_3 x + a_4 = 0,$$

where the coefficients are:

$$\begin{aligned} a_1 &= 4\beta_1\beta_2, \\ a_2 &= 3\beta_1 - 3\beta_2 - 6\beta_1\beta_2 - 6s_1\beta_1\beta_2 + 6s_2\beta_1\beta_2, \\ a_3 &= -2 - 2\beta_1 - 4s_1\beta_1 + 4\beta_2 - 4s_2\beta_2 + 2\beta_1\beta_2 + 8s_1\beta_1\beta_2 + 2(s_1)^2\beta_1\beta_2 \\ &\quad - 4s_2\beta_1\beta_2 - 8s_1 s_2\beta_1\beta_2 + 2(s_2)^2\beta_1\beta_2, \\ a_4 &= 1 + 2s_1\beta_1 + (s_1)^2\beta_1 - \beta_2 + 2s_2\beta_2 - (s_2)^2\beta_2 - 2s_1\beta_1\beta_2 - 2(s_1)^2\beta_1\beta_2 \\ &\quad + 4s_1 s_2\beta_1\beta_2 + 2(s_1)^2 s_2\beta_1\beta_2 - 2s_1(s_2)^2\beta_1\beta_2. \end{aligned}$$

Define the following relationships:

$$Q \equiv \frac{a_3}{3a_1} - \left(\frac{a_2}{3a_1}\right)^2,$$

$$R \equiv \frac{a_3 a_2}{6(a_1)^2} - \frac{a_4}{2a_1} - \left(\frac{a_2}{3a_1}\right)^3,$$

$$D \equiv Q^3 + R^2,$$

$$S \equiv \left(R + \sqrt{D}\right)^{\frac{1}{3}},$$

$$T \equiv \left(R - \sqrt{D}\right)^{\frac{1}{3}}.$$

D is the discriminant that determines the nature of the roots of the equation. If $D > 0$, there are one real root and two conjugate complex roots; if $D = 0$, there are real roots of which at least two are equal; if $D < 0$, there are three distinct real roots. In this model, D is negative and there are three distinct real roots. Cardano's formulae for the roots are as follows:

$$root_1 = -\frac{a_2}{3a_1} + (S+T),$$

$$root_2 = -\frac{a_2}{3a_1} - \frac{1}{2}(S+T) + \frac{1}{2}i\sqrt{3}(S-T),$$

$$root_3 = -\frac{a_2}{3a_1} - \frac{1}{2}(S+T) - \frac{1}{2}i\sqrt{3}(S-T),$$

where $i = \sqrt{-1}$. For this model, it turns out that $root_2 < root_3 < root_1$. The solution is only defined for $\beta_1 > 0$ and $\beta_2 > 0$. To make sure the solution is the Nash bargaining solution, check that $(u_1(x_1^N), u_2(1 - x_1^N)) > (0,0)$. The optimal solution is $x_1^N = root_3$, and the Nash bargaining solution is $(x_1^N, 1 - x_1^N)$.

The Nash bargaining solution: A special case

We here provide the Nash bargaining solution for the special case where one individual is fairness motivated and the other is only motivated by material self-interest, as referred to in footnote 15. We assume:

$$u_1 = x_1 - \beta_1(x_1 - s_1)^2,$$

$$u_2 = x_2,$$

$$u_i(d, s_i, \beta_i) = 0.$$

By substituting the constraint into the objective function, the optimization problem can be written as:

$$\max f(x_1) = \left(x_1 - \beta_1(x_1 - s_1)^2\right)\left(1 - x_1\right).$$

Differentiating f with respect to x_1 gives a quadratic equation, where the root that satisfies the Nash bargaining solution is,

$$x_1^N = \frac{1 + \beta_1 + 2\beta_1 s_1 - \sqrt{1 - \beta_1 + 4\beta_1 s_1 + (\beta_1)^2 - 2(\beta_{1)}^2 s_1 + (\beta_1 s_1)^2}}{3\beta_1}.$$

The Nash bargaining solution has the property that it approaches the solution for two self-interested individuals when fairness motivation is weak, $\lim_{\beta_1 \to 0}(x_1)^N = \frac{1}{2}$. For a strongly fairness motivated individual the solution approaches his fair share, $\lim_{\beta_1 \to \infty}(x_1)^N = s_1$.

The effect of a change in β_1 can be derived by differentiating the Nash bargaining solution for this special case,

$$\frac{dx_1^N}{d\beta_1} = \frac{2 + \beta_1(4s_1 - 1) - 2\sqrt{1 + \beta_1(4s_1 - 1 + \beta_1(s_1 - 1)^2}}{6(\beta_1)^2\sqrt{1 + \beta_1(4s_1 - 1 + \beta_1(s_1 - 1)^2}}.$$

This expression is zero for $s_1 = 0.5$, it is positive for any change in β_1 provided that $s_1 > 0.5$, and it is negative if $s_1 < 0.5$. Hence, when the fairness motivation of an individual increases, the Nash bargaining solution assigns a share that is closer to his fair share.

Appendix B Experiment instructions

General introduction

Welcome to this experiment. My name is (...) and I will guide you through the experiment. The results from the experiment will be used in a research project, and it is therefore important that you all stick to the rules that have been distributed:

– You should not talk to other participants.
– If you have questions or problems during the experiment, raise your hand and we will come to you.
– You should not open other web pages.

If you breach these rules, you will have to leave the room. There will be pauses during the experiment and it is important that you sit still and keep quiet during these.

You will be completely anonymous in the experiment. You will not at any time be asked about who you are. It will not be possible for us or the other participants to find out which choices you have made. You will be asked to make choices in several different situations in this experiment. For every situation, you will be randomly connected to another person in this room. Your actual payment will be determined as follows: we randomly draw one of the situations you were involved in and pay the amount of money you received in that situation. The choices that you make will not influence which situation is drawn; it will be an entirely random draw and there is an equal chance for all situations to be drawn. You should therefore think about each situation as if it is the one that determines how much you will earn.

When the experiment is finished, you will see a payment code on the screen. You are asked to write down this code on a form that will then be sent to the accounting department at (...). Employees at the accounting department will receive a list of codes and amounts from us and match these with the payment instructions from the forms. This is done so that nobody will know how much you have earned.

The experiment consists of four phases. I will now explain the main features of the experiment. I will stop before we start a new phase and explain in more detail what you should do in each phase. In the first phase of the experiment, you will be copying text in Word for 10 minutes . You will be paid a price for each correct word you have typed. In phase two of the experiment you will be randomly matched with other persons in this room, and each person in each pair will choose how much of the combined production value to distribute to yourself and to the other person. You will be involved in four such *situations of distribution*.

In the third phase of the experiment, you will also be randomly matched with people in this room. You will then *negotiate* about the division of the combined production value by sending proposals to each other until one of you accepts the other's proposal. The production value shrinks by 4% every time one of you does not accept the other's proposal. You will be involved in four such *situations of negotiation*. In the last phase of the experiment you will be asked to answer a few questions about the types of situations that you have experienced.

Introduction phase 1

The first thing you will do is to copy text from an official report that is marked with either an A or a B, and which you will find in the folder on your desk. You will start copying the text into Word when I tell you. When 10 minutes have passed, I will let you know and everybody must then stop. You will be paid for each correct word you type. You may use the spellchecker in Word.

I remind you that you should raise your hand if you have any problems or questions, and then someone from the research group will come and help you. You can now open a new document in Word and we will soon start to type. *You can start typing now* .

Everybody must now stop typing. You should now highlight all the text typed and copy it to the window in the Mozilla browser, then click on the button marked 'submit text'.

After having submitted the text you will see a screen that shows how much you have produced and the value of your production. The production is rounded off to the nearest 50 words. Half of you have copied text marked A, which is an excerpt from an official report on the merger of the telecoms, IT, and media sectors. You will receive one krone and 50 oere for each correct word you have typed. The other half has copied text marked B, which is an excerpt from an official report about Norwegian performing art. You will receive 75 oere for each correct word you have typed. These prices are randomly determined by us. Finally, click on the button marked 'continue'.

Introduction phase 2

You will now be randomly matched with other people in this room. In each *situation of distribution* you will not know who the other person is, and the other person will not know who you are. You will be informed about how many words he or she has produced and what price each of you has randomly been allocated. You will then choose a *distribution* of the combined production value between you and the other person. Remember that this is real money and that the way that you divide the money determines how much you earn and how much the other person earns. You will be asked to make decisions in two such *situations of distribution*. In two other situations of distribution, another person will decide how much he or she will distribute to you.

After you have registered the distribution, you will see a new screen where you are asked either to confirm the distribution or to go back and change the distribution. When you have confirmed your choices, you will receive a message that you have finished the second phase of the experiment. You should then quietly wait for all the other people in the room to finish making choices in their situations. On the computer you will soon see a screen with the first situation and you can then start making choices.

Introduction phase 3

Everybody has finished the second phase and I shall now explain what you will be doing in the third phase of the experiment. You will this time also be randomly matched with other people in this room. In each *situation of negotiation* you will not know who the other person is and the other person will not know who you are. You will be informed about how many words the other person has produced and what price he or she has randomly been allocated. One of you is randomly drawn to make the first proposal for division of your combined production value. The proposal will be sent to the other person and he or she has two choices: to accept your proposal or to make a new proposal for division. New proposals are sent back and forth until one of you chooses to accept the other‘s proposal. Every time one of you does not accept the proposal for division, but comes up with a new proposal, the remaining production value will be reduced by 4%. Everybody will be involved in four such negotiation situations.

In some situations you will be asked what you think will be the final outcome of the negotiation. If your answer is within a deviation of plus or minus 20 kroner of the actual result, you will receive 20 kroner in extra payment, with one exception: if you guessed the negotiation result in a situation, and this particular result was randomly drawn, you will *not* receive the extra payment for a correct guess but only the payout in this situation. You will also be asked to state how *certain* you are about your guess. Your answer should be given in terms of a certainty percentage, that is, a number between 0 and 100. It is important that you write a high percentage if you are certain that this will be the final result, and a low percentage if you are uncertain if this will be the final result.

On the computer you will soon see a new screen with the first situation and you can then start to negotiate. When the situation is accepted, you will automatically get a new situation to negotiate. When you have finished all the negotiation situations, you will be asked to wait until everybody has finished their choices.

Introduction phase 4

Everybody has finished and we will soon draw the situation that will decide your payment from this experiment. First, we ask you to answer a few questions. Soon you will see a new screen with information about the first question. Click on the button marked ‘go forward’ when you have read the information and thereafter please answer all the questions.

Closing and payment

Everybody has now answered the questions. You will soon see a screen that informs you about which situation that has been drawn randomly, and how much you have earned in this situation. This screen will be open for 45 seconds. Thereafter you will automatically be forwarded to a new screen, which only contains a payment code.

Everybody now has a screen with the payment code. Write down this payment code on the form that you find in the folder next to you. On the form also write your name and bank account details. Put the form in the envelope and place the envelope in the box by the door when you leave the room.

The experiment is now finished and, on behalf of the research team, I thank you again for your participation in this experiment.