Paper submission for the Jahrbuch für Handlungs- und Entscheidungstheorie Bd. 10, 2016

Mitigating the Problem of Manipulation in the 'Adjusted Winner' Procedure

The 'Adjusted Winner' procedure (AW) is a mechanism to reach fair agreements in bargaining situations over a fixed set of objects. A major shortfall of AW for both mediators and participants is that it relies on participants' honesty, which makes it open for manipulation. In a general model of AW for two objects and continuous manipulation strategies, it is shown that a) manipulation is always risky since potential losses are always larger than potential gains; and b) there exists an equilibrium of symmetric manipulation and with equal threat of potential losses that leads to exactly the same outcome as truthful behavior. These findings imply that the problem of manipulation in AW is mitigated.

Keywords: Adjusted Winner, Manipulation, Bargaining, Negotiation, Fairness

1. Introduction

Bargaining is a ubiquitous feature of social interaction. Politicians negotiate agreements frequently – be it the post-election bargaining over offices or the settlement of international disputes. In people's daily lives, bargaining and conflict play an equally important role: Take for example haggling at a car dealership or pay-raise negotiations with an employer.

A central issue in bargaining is the question of fairness. Fairness in general has stirred theoretical-philosophical enquiry across many disciplines; from Plato's formula 'that each should get what he deserves' to John Rawls' (1985) 'Justice as Fairness'. In addition, economists like John F. Nash (1950) tried to find a rational and, as he claimed: A solution fair solution in bargaining situations (Nash 1950, p. 158).

Furthermore, evidence from psychological experiments suggests that some people exhibit a preference for fairness. As outlined for example by Fehr and Schmidt (1990), people are willing to sacrifice their own gains for the sake of the gains of others in simple ultimatum bargaining experiments.

The question of fairness in bargaining is mostly asked from an *outside* point of view, which asks: Given the characteristics of the situation and the players, what can be considered a just distribution according to external principles? In contrast, individuals *within* such situations are often assumed to only care about benefitting individually. While it seems that these questions are separate from each other, one main point of this paper is to argue that – especially for practical purposes – they should be thought about together.

Evidence from psychological experiments suggests that some people exhibit a preference for fairness. As outlined for example by Fehr and Schmidt (1990), a small number people are willing to sacrifice their own gains for the sake of the gains of others in simple ultimatum bargaining experiments. The benefit of posing the additional question of incentive compatibility is that, instead of trusting people's adherence to fair solutions, distributional mechanisms in which individual incentives lead to fair outcomes can be found. I discuss this problem of incentive-proof fair division for the 'Adjusted Winner' (AW) procedure by Brams and Taylor (1999) because AW is a practically interesting fair division procedure and a vivid example of the described dilemma.

In part two, I describe how the AW procedure works, why it can be said to lead to a fair outcome and give an example of a practical application of the method. I go on to argue that while practical issues limit the scope of AW's applicability, the problem of manipulation poses a more severe and even fundamental threat to it: If players manipulate, the fairness properties of the AW solution can no longer be guaranteed.

The main contribution of this paper, located in part three, is the description of a model that envisages AW as a strategic game with continuous manipulation strategies for the case of distributing two objects. Through the introduction of continuous strategies, the model in this paper fundamentally diverges from previous approaches, such as Schüssler's (2007) who discusses a situation with a binary strategy choice.

The model yields two remarkable findings that mitigate the problem of manipulation in AW. *First*, the model shows that the maximal potential gains from manipulation are always smaller than the minimal losses in case manipulation fails. This implies that a manipulative strategy is dangerous for a manipulating player if she is not entirely sure about her opponent's valuation or strategy.

Second, if both players simultaneously announce their manipulated preferences, a Nash equilibrium exists which results in the same payoffs as in the case where both players announce their valuations truthfully. This specific Nash equilibrium is an appealing solution because it is symmetric and both players face the same potential loss through deviation from their strategy. Furthermore, as long as both manipulate equally strongly and do not change the initial distribution of objects, the appealing properties of the AW solution are preserved. This suggests that the problem of manipulation in AW is mitigated. To show this, I start by outlining the AW procedure in the next section.

2. The AW Procedure

2.1 How does AW work?

Adjusted Winner can be employed to allocate a fixed number of objects between two[1] players fairly. The objects at stake are assumed to be arbitrarily divisible and linear in their utility when divided. Also, the utility of having one object is independent from having any other objects. These requirements are admittedly strong and lead to rather strict limitations for AW, which will be further discussed in light of the example of the Camp David negotiations below.

AW proceeds in three steps. The first step for the players is to assign a total of 100 points to the objects at stake, whereas their allocation of points must reflect their preferences about these objects. Through that step, AW surveys and normalizes the players' preferences in an easy and

[1] Note that the generalization to n players has consequences for AW. The procedure becomes more complicated and loses some of its appealing properties.

understandable way. However, note that this step requires a cardinal interpretation of utility: If a solely ordinal utility scale was assumed, any allocation of points that puts the objects in the same order would be indistinguishable.

In the second step, the distribution of the objects is carried out: Each object goes to the player who gave the object more points in the first step. In the case where an object has received equal points from both players, the object goes to the player who received fewer points so far.

In a last step, the gains from step two are 'adjusted': The object which is most similar in valuation is partly transferred from the 'richer' player to the 'poorer' player until both are in the possession of (shares of) objects to which they have assigned the same amount of points.

The term 'most similar in valuation' refers to the ratio of the players' point allocations to the same objects *O*, i.e. $r_o = \frac{v_{rich}}{v_{poor}}$. The closer this ration is to one, the more similar in valuation is the object for the two players. Of course, only objects with $v_{rich} > v_{poor}$ can be redistributed. If one object's complete redistribution does not suffice to equal out the player's points, the procedure continues analogously with the good which is now the most similar in valuation and still belonging to the richer player. In summary, the three steps are as follows:

> *Step 1:* Both players allocate 100 points truthfully to the objects at stake.
>
> *Step 2:* Each object goes to the player who allocated more points to it. Objects with equal points go to the player who received fewer points in total so far.
>
> *Step 3:* The object for which the valuation is most similar is partially redistributed so that both receive objects worth the same amount of points.

A simple example can illustrate the steps of AW (see Brams and Taylor 1999, p. 72). Consider two players facing the task of distributing five not further specified objects A, B, C, D and E. In *step 1*, the players allocate their 100 points as shown in Table 1:

Table 1: An exemplary point allocation for AW

Object	**Player 1**	**Player 2**
A	<u>50</u>	40
B	20	<u>30</u>
C	<u>15</u>	10
D	10	10
E	5	<u>10</u>

The higher valuation for each object is underlined. In *step 2*, player 1 receives objects A and C, while player 2 receives B and E. Object D also goes to player 2, because she has only received objects worth 40 points to her, while player 1 has received 65 points.

Step 3: Player 1 is richer (65 points vs. 50 points so far). The object closest in valuation that has been given to player 1 is A with a ratio of $r_A = 1.2$, while for object C, $r_C = 1.5$. Therefore, A must be redistributed so that they have objects worth equal points to them afterwards.

To calculate how much of A must change hands, let *x* be the percentage of object A that will be transferred from player 1 to player 2. The condition that player 1 and 2 must have objects of the same point-value after the transfer translates into the following equation:

$$u_1(C) + (1 - x) * u_1(A) = u_2(B\&D\&E) + \ x * u_2(A)$$

This leads to the following equation via plugging in the actual numbers:

$$15 + (1 - x) * 50 = 50 + x * 40.$$

Solving for x yields $x = 1/6$. This means that player 1 has to give $1/6$ of A to player 2. The final allocation is: Object C and $5/6$ of A go to player 1; B, D, E and $1/6$ of A go to player 2. Each player therefore receives objects worth 56.67 points to her.

2.2 Why should the outcome of an AW process be considered fair?

Raith (2000) shows that AW yields the Kalai-Smorodinsky Bargaining Solution (KSBS). The KSBS is the only solution that fulfills a certain set of axiomatic requirements (see Kalai and Smorodinsky, 1975). That AW or the KSBS carries a central notion of fairness is expressed in the following axioms.

The *symmetry* axiom demands that indistinguishable players receive the same payoffs. Or put shortly: Equals should be treated equally. The *monotonicity* axiom accounts for how to treat players with different preconditions. It requires that a player should not be worse off if the space of feasible solutions is increased in her favor, or intuitively: ‘new options for a player should never be a disadvantage’.

The KSBS can be derived as the unique solution when *symmetry*, *monotonicity* as well as *Pareto-efficiency* and *positive linear transformability* of utility functions are required as further axioms. In the KSBS, each player has her best outcome satisfied to the same extent, and it is the solution that benefits each player the most.

Since AW yields the only solution that embodies these axioms of fairness all at the same time, it can be considered a fair mechanism that results in a fair outcome. Certainly, one can disagree with the particular notion of fairness employed here, and how it is captured in the axioms. Nonetheless, it is a reasonable and analytically sound argument that formalizes those principles and gives a clear derivation of their implication.

Those abstract properties translate into the practical merits of envy-freeness, efficiency and equitability (Brams and Taylor 1999 p. 69). AW's solution is *envy-free* because in the resulting allocation, neither player would want to exchange her final bundle of objects with the other player. It is *efficient* because no Pareto-improvements are possible in the final allocation. It is *equitable* (or fair) in the sense of the KSBS, that each player's optimal preference is realized to the same degree. Based on this theoretical argument, it is reasonable to adhere to the AW solution as a fair outcome.

2.3 An example: AW and the Camp David Accords

Brams and Taylor (1999, p. 69) also argue that AW can be performed in a relatively simple way. While this may be true for the steps of the mechanism itself, the range of applicability is a crucial challenge for AW. Doubtlessly, there are severe limitations that come along with its requirements. Divisibility, linear and independent utility are met only by very few objects. However, even though this may render AW inapplicable in a variety of situations, there are still cases where those requirements are at least reasonable approximations of reality.

The example of the negotiations between Egypt and Israel, which took place in Camp David, USA, in 1978 illustrates this. The Camp David talks were the conclusion of a long process of peace talks after several violent conflicts between the two countries. After 13 days of negotiations, leaders of both parties finally signed an agreement, later to be known as the 'Camp David Accords'.

Brams and Taylor (1999, p. 89ff) use AW to assess whether the outcome of the Camp David negotiations was fair for both parties. They identify six major issues in the process and reconstruct the Israeli and Egyptian preferences based on expert judgments. Those are described in Table 2 below.

Table 2: Issues and implied point allocation in the Camp David negotiations

Issue	Israel	Egypt
Sinai	35	<u>55</u>
Diplomatic recognition	<u>10</u>	5
West Bank / Gaza strip	<u>20</u>	10
Linkage	<u>10</u>	5
Palestinian rights	5	<u>20</u>
Jerusalem	<u>20</u>	5

Issues 1, 3 and 6 stand for having control over the respective area. Issue 2 stands for *diplomatic recognition* of Israel, which Israel favored and Egypt opposed. Equally, *Palestinian rights* were advocated by Egypt, and not recognized by Israel. *Linkage* embodies Egypt's claim that the success of the negotiations at hand must be formally linked to the progress of recognition of Palestinian autonomy. In these three latter cases, 'winning' would mean to get one's way in the decision (see Brams and Taylor 1999, p. 91ff for a more elaborate description and discussion of the issues).

AW's solution would prescribe that issues 1 and 5 go to Egypt, while issues 2, 3, 4 and 6 are ascribed to Israel. Then, to even out point gains, 1/6 of the issue 'Sinai' must be redistributed to Israel. With the final distribution, each player receives objects worth 66.7 points.

Brams and Taylor argue that the actual outcome after the talks closely resembles the solution that is prescribed by AW. All issues were allocated accordingly; the division of the issue 'Sinai' was accomplished through the following compromise: Israeli military bases and civil settlements were removed, but the Sinai Peninsula was demilitarized and U.S. troops were stationed to monitor the enforcement of the agreement. According to Brams and Taylor (1999., p. 97), this can be envisaged as a redistribution of 1/6 of the issue. Thus, they argue, the analysis using AW allows to qualify the Camp David agreement as fair.

Note that Brams and Taylor employ AW in a manner of a ‘hypothetical procedure’. AW was not actually used during the Camp David talks. The actual outcome from Camp David is merely compared to the fair outcome that AW would have prescribed.

2.4 General Problems for the Application of AW

There are several practical issues that could be discussed at this point: Are the identified issues really independent from one another? Can the agreement on the Sinai issue be considered a 1/6th- split? Could the other issues have been split as well if the procedure had required it? How reliable is the point allocation based on expert judgments?

Certainly, if one wants to apply AW as an actual negotiation tool, it is always problematic whether or not those conditions are fulfilled. Certainly, there will be cases where the issues at stake do not allow the application of AW because dependencies and non-linearities are too strong. Yet, there are cases that meet the conditions of AW in reasonable approximation as in the Camp David case. One therefore needs to be aware of those issues, be able to handle them or use AW only in situations where they apply sufficiently well.

However, there is a more severe problem for the application of AW as an actual negotiation tool. This is the problem of manipulation. Given that the involved actors know how the mechanism works, they should be assumed to be trying to shift outcomes in their favor. Therefore, to convince actors of the usefulness of AW, one should appeal to individual incentives rather than benevolence or a desire for fairness. Instead of a cooperative mechanism, AW then becomes a strategic game between rational actors. AW (like other fair division mechanisms) must be able to work under the assumption of selfish utility maximizers. If actors were cooperative anyways, there would be no need for a dispute settling mechanism.

2.5 Optimal Manipulation in the Camp David Example

Schüssler (2007, p. 290f) scrutinizes the Camp David example and assesses strategies of manipulation in the case where the true valuations are known by the other side. An intuitively optimal manipulation strategy, given that the opponent announces truthful valuations, is to win the same issues as in AW, but to win each issue only by a slight margin. One can then allocate the 'saved' points to the issues the manipulating player loses. This leads to either a higher share being redistributed to the manipulating player, or the manipulating player has to give up less of her goods through redistribution.

For instance, following the above strategy intuition most closely, Israel could announce the point scheme 53, 6, 11, 6, 18, 6 in the Camp David example, as depicted in Table 3 below.

Table 3: Truthful and manipulated point allocations in the Camp David example

Issue	Israel (true)	Israel (manipulated)	Egypt (true)
Sinai	35	53	55
Diplomatic recognition	10	6	5
West Bank / Gaza strip	20	11	10
Linkage	10	6	5
Palestinian rights	5	18	20
Jerusalem	20	6	5

In that case, the first round distribution of AW would remain the same as before. However now, Israel receives issues worth only 27 points (according to the manipulated point allocation), while Egypt still gains 75 first- round points. Through the AW redistribution mechanism, 12/27 of Sinai must now be given to Israel. Israel's payoff (for which its true valuation from the table above is used) is then a satisfaction of interests worth 75 points compared to 66.7 points for truth telling. Egypt's satisfaction would be reduced to 51 points through Israel's manipulation.

Schüssler shows that with a similar strategy, Egypt could even gain 79 points and reduce truth-telling Israel's satisfaction to 51 points. Yet, if both parties announce their manipulated valuations, both would receive only 34 points: The efficient allocation of objects would be reversed and each would get those objects she likes less.

Given a binary choice between manipulating (in this particular way) or being truthful, this would leave the players in a 'chicken game': Whoever convinces the other that she will manipulate will force the other to comply by stating her true preferences (since this is the best response to the described manipulation strategy). There are two pure strategy equilibria (one manipulator, one truth-teller) and one mixed-strategy Nash equilibrium. For applications of AW, this would be a problem because it is not clear how players would act, and manipulation must be expected to occur at least with a certain probability. The situation is depicted in Table 4 below in normal form (see also Schüssler 2007., p. 290).

Table 4: AW as a strategic game in the Camp David example with binary strategy choice

		Israel	
		truth-telling	**manipulate**
Egypt	**truth-telling**	66.7 / 66.7	51 / 71
	manipulate	79 / 51	34 / 34

Schüssler's description simplifies the situation in one central regard: It models the actors' strategy choice as binary. One can either manipulate (in an optimal way, given that the other player is truthful) or be truthful. This reduces the space of options considerably. Given how AW functions, each player can announce any valuation she wants. Hence, also manipulating 'a little bit' is feasible. In the following section, I account for this possibility of continuous manipulation for both players. However, this modeling approach limits itself to a generic case with only two goods. Generalizing those results to the case of *n* goods is far from straight forward and involves manifold mathematical complications. Thus, the analysis here is limited for the sake of clarity and solvability in order to get first insights before tackling the more complicated *n* items case.

3. A General Model for Manipulation in AW

This section develops a model of manipulation in AW with continuous strategies for the case of two objects. In the model, both players can choose to announce any valuation for the objects at stake. The model shows that the threat of manipulation is mitigated because manipulation is risky and mutual manipulation can cancel out.

Due to the complexity of the mathematical analysis, all results are derived for the case of two players (1 and 2) with only two objects (A and B). This, of course, has an impact on the generalizability of the findings, which will be discussed afterwards. Further, all the practical requirements for AW will be assumed to be fulfilled. This means that A and B's utility are linear for shares of objects, and the utility from having a share of A is independent from one's share of B (and vice versa). While the discussion of those aspects is also of high relevance, they are more suited for empirical analysis rather than an abstract-mathematical approach as the one taken here; this paper focusses on the threat through manipulation.

3.1 Setup and Payoff Functions

Let $\widetilde{v_i}$ denote player i's true valuation for object A (i.e. the percentage of points she would truthfully allocate to object A in AW). It is assumed that $0 < \widetilde{v_i} < 1$, which means that both players have at least some interest in each individual good. It follows that player 1's valuation for object B must be $1 - \widetilde{v_i}$. The same is true for player 2. Also, $\widetilde{v_i}$ is a continuous variable, so that any real fraction of whole points can be allocated. Table 5 below depicts the notations for the true valuations.

Table 5: Notation for both player's true valuations

	Player 1	**Player 2**
Object A	$\widetilde{v_1}$	$\widetilde{v_2}$
Object B	$1 - \widetilde{v_1}$	$1 - \widetilde{v_2}$

Further, assume $\widetilde{v_1} > (1 - \widetilde{v_1})$ without loss of generality. This is just the convention to choose object A to be the object that player 1 likes more.

Let v_i denote the valuation which player i *announces*. This is the value that affects the distribution from AW. Further, $v_i \in [0,1]\ \forall i$ and v_i is also continuous. For a truthful player, the announced valuation is equal to her real valuation, hence $\widetilde{v_i} = v_i$. Both players can also choose to manipulate by announcing a valuation $v_i \neq \widetilde{v_i}$. Again, it is assumed that all points are allocated, meaning that player i announces v_i for object A, and $1 - v_i$ for object B.

The first important step is identifying the player's payoff functions. The following section will reconstruct this for player 1, since this is the point of view taken henceforth. Player 2's payoff function can be obtained simply by switching indices.

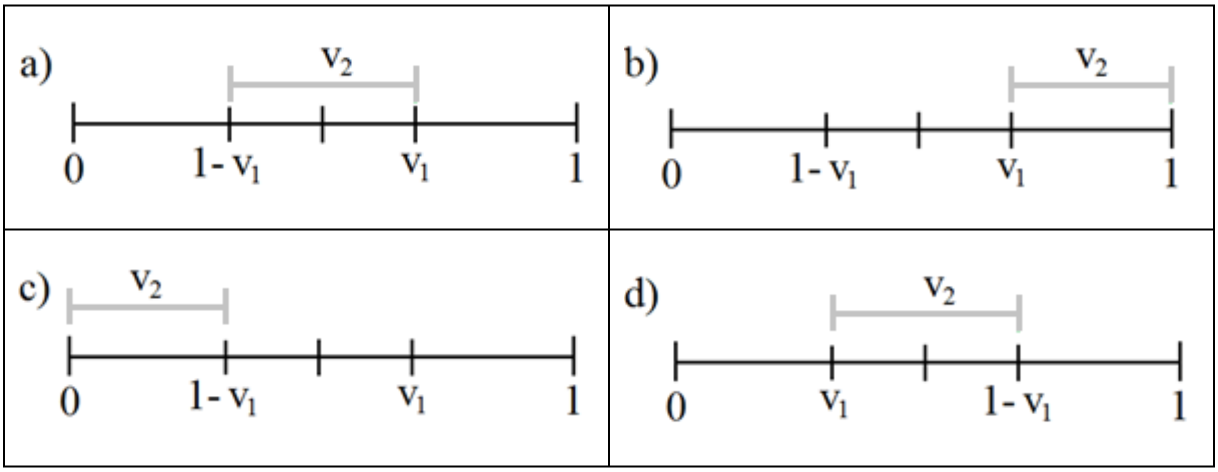
Player 1's payoff function depends on her true valuation $\widetilde{v_1}$, which is a not further specified random parameter. The payoff function is further determined by both players' strategic announcements v_1 and v_2, which they can choose freely. Two factors must be distinguished for calculating the payoff function:

First, did player 1 announce a higher valuation for object A than player 2 did? This question decides whether player 1 receives object A or object B initially. In mathematical terms, is $v_1 > v_2$? If yes, she will be given object A in step 2 of AW. If not, she gets object B.

The second question is: Which player realized more points in step 2 of AW? This question decides whether player 1 has to give something to player 2, or vice versa. For instance, if $v_1 > v_2$ and therefore player 1 received A and player 2 received B, the question is: Is $v_1 < (1 - v_2)$, or the other way around? The player who received more will have to give up a share of her object in order to equal out points. In total, this results in four cases that stem from these two questions. Fig. 1 visualizes those four cases. Those cases can be described in colloquial language as follows:

a) Player 1 receives A and gives parts of A to player 2
b) Player 1 receives A and receives parts of B from player 2
c) Player 1 receives B and gives parts of B to player 2
d) Player 1 receives B and receives parts of A from player 2

Fig. 1: Visualization of all possible constellations of v_1 *and* v_2



Note that even though it is assumed that $\widetilde{v_1} > 1 - \widetilde{v_1}$, this must not necessary hold for v_1. Player 1 can very well chose to announce a valuation below 0.5 for object A, even if her true valuation for A is assumed to be larger than 0.5.

Consider case (a) where $v_1 > v_2$ and $v_1 > 1 - v_2$. In the initial allocation of objects, player 1 receives object A; player 2 receives object B. Since by assumption $v_1 > 1 - v_2$, player 1 realized more points in step 2, hence x percent of good A must be redistributed from 1 to 2.

The relevant condition for the calculation of x is that both players have equal overall points after redistribution. Player 1 gives up x percent of A, while Player 2 receives x percent of A. Note that for the calculation of x, only the *announced* valuations are relevant. This leads to the following equation:

$$(1 - x) * v_1 = (1 - v_2) + x * v_2$$

$$x = \frac{v_1 + v_2 - 1}{v_1 + v_2} = 1 - \frac{1}{v_1 + v_2}$$

For the calculation of the payoff, the *true* valuations must be employed since these are the values that determine a player's actual payoff. In the final allocation, player 1 holds $(1 - x)$ of good A, therefore her payoff is

$$P_1^{(a)} = (1 - x) * \widetilde{v_1}$$

$$P_1^{(a)} = \left(1 - (1 - \frac{1}{v_1+v_2})\right) * \widetilde{v_1}$$

$$P_1^{(a)} = \frac{\widetilde{v_1}}{v_1+v_2}$$

The calculations for the other three cases run along similar lines and are given in appendix 1. Table 6 below gives the payoff function for player 1 for all four cases. Note that all P_i's are a function of variables v_1 and v_2, hence $P_i(v_1, v_2)$. This is omitted for the sake of brevity and will be denoted only as P_i.

Table 6: Payoff functions for player 1

	$\boldsymbol{v_1 > v_2}$	$\boldsymbol{v_1 < v_2}$
$\boldsymbol{1 - v_1 < v_2}$ $\boldsymbol{v_1 > 1 - v_2}$	(a) $P_1^{(a)} = \frac{\widetilde{v_1}}{v_1+v_2}$	(b) $P_1^{(b)} = 1 - \frac{\widetilde{v_1}}{v_1+v_2}$
$\boldsymbol{1 - v_1 > v_2}$ $\boldsymbol{v_1 < 1 - v_2}$	(c) $P_1^{(c)} = 1 - \frac{(1-\widetilde{v_1})}{(1-v_1)+(1-v_2)}$	(d) $P_1^{(d)} = \frac{1-\widetilde{v_1}}{(1-v_1)+(1-v_2)}$

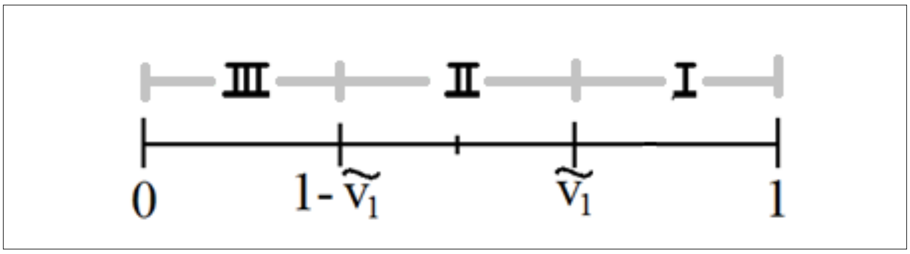
3.2 Optimal Manipulation against a Truthful Player

Now assume that player 2 will always announce her true valuation, i.e. $v_2 = \widetilde{v_2}$. First, the optimal manipulation strategy for player 1 in this case is discussed. It can be generally shown that player 1's optimal manipulation strategy is to let v_1 approach $\widetilde{v_2}$, thus proving for two objects what Schüssler (2007) described on an intuitive level for more than two objects. However, as the subsequent section shows, this strategy is very dangerous because the maximal gains from optimal manipulation (compared to an honest strategy) is always smaller than the loss player 1 would suffer from only the slightest over-manipulation.

Technical hint: It is implicitly assumed that, if v_1 is exactly on the border of two payoff functions, player 1 can 'choose' which of the two bordering payoff functions is used. The correct notation for this would be for instance $v_1 = \lim_{\varepsilon \to 0}(\widetilde{v_2} - \varepsilon)$ for $\varepsilon > 0$ to show that v_1 approaches $\widetilde{v_2}$ from below, so that still the payoff function for the case $v_1 < \widetilde{v_2}$ is applicable. This detailed notation is omitted here for the sake of brevity. Originally, AW prescribes that in case of equal points, the object goes to the player with fewer total points, and is then potentially redistributed. The problem of 'choosing a payoff function' does not occur in the original AW-procedure, since payoffs are the same under both functions.

Under the assumption $\widetilde{v_1} > (1 - \widetilde{v_1})$, three different constellations for $\widetilde{v_1}$ and $\widetilde{v_2}$ can occur. $\widetilde{v_2}$ can be either in region I, II or III in relation to $\widetilde{v_1}$, as depicted in Fig. 2 below.

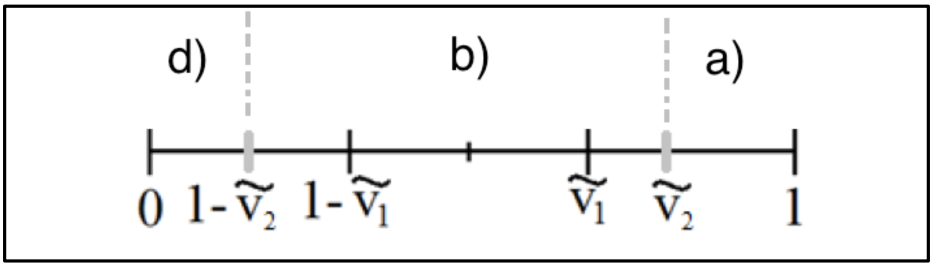
Fig. 2: Three possible constellations of true valuations



In the following, the optimal manipulation strategy for constellation I will be shown. The same results can be obtained for constellations II and III, for which the calculations are in appendix 2.

For this part, assume therefore that $\widetilde{v_2}$ lies in area I, and hence $1 - \widetilde{v_1} < \widetilde{v_1} < \widetilde{v_2}$. Depending on player 1's choice of v_1, either payoff function (a), (b), or (d) is applicable. *Fig. 3* below depicts the regions in which the different payoff-functions apply.

Fig. 3: Applicable payoff function depending on choice of v_1



- Player 1's payoff function is $P_1^{(a)} = \frac{\widetilde{v_1}}{v_1 + \widetilde{v_2}}$ for $v_1 > \widetilde{v_2}$ (since it follows $v_1 > 1 - \widetilde{v_2}$) Function (a) strictly decreases with v_1, hence the local maximum is reached for $v_1 = \widetilde{v_2}$. The maximum payoff is then $P_1^{(a)max} = \frac{\widetilde{v_1}}{2\,\widetilde{v_2}}$.

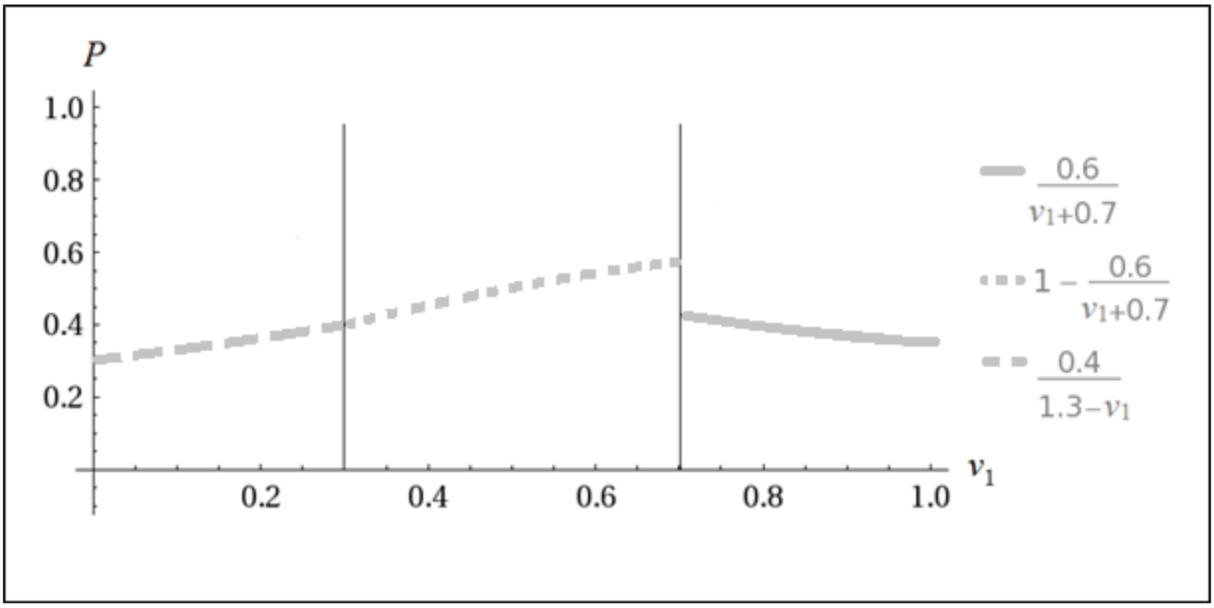
- Player 1's payoff function is $P_1^{(b)} = 1 - \frac{\widetilde{v_1}}{v_1 + \widetilde{v_2}}$ for $1 - \widetilde{v_2} < v_1 < \widetilde{v_2}$ (since it follows $1 - v_1 < \widetilde{v_2}$). Payoff function (b) strictly increases with v_1, hence the local maximum is reached for $v_1 = \widetilde{v_2}$, which is the largest value of $v_1$in (b). The maximum payoff is then $P_1^{(b)max} = 1 - \frac{\widetilde{v_1}}{2\,\widetilde{v_2}}$.

- Player 1's payoff function is $P_1^{(d)} = \frac{(1-\widetilde{v_1})}{(1-v_1)+(1-\widetilde{v_2})}$ for $v_1 < 1 - \widetilde{v_2}$ (since it follows $1 - v_1 > \widetilde{v_2}$). Payoff function (d) strictly increases with v_1, hence the local maximum is reached for $v_1 = 1 - \widetilde{v_2}$, again the largest value of $v_1$in (d). The maximum payoff is then $P_1^{(d)max} = (1 - \widetilde{v_1})$.

The local maxima can now be compared. Since we are still in case I where $\widetilde{v_1} < \widetilde{v_2}$, it follows that $P_1^{(b)max} > P_1^{(a)max}$. Therefore, player 1 prefers being in (b) over being in (a) in constellation I. Also, $P_1^{(b)max} > P_1^{(d)max}$, which means that player 1 also prefers (b) over (d).

Looking at the complete payoff function, the global maximum is $P_1^{max} = P_1^{(b)max}$. Therefore, player 1's optimal strategy is to chose $v_1 = \widetilde{v_2}$ but still stay in (b) and hence a little below $\widetilde{v_2}$ (technically correct: $v_1 = \widetilde{v_2} - \varepsilon$ with $\varepsilon \to 0$ and $\varepsilon > 0$). Fig. 4 illustrates the payoff function for an example of constellation I with $\widetilde{v_1} = 0.6$ and $\widetilde{v_2} = 0.7$.

Fig. 4: Exemplary payoff function for constellation I with $\widetilde{v_1} = 0.6$ *and* $\widetilde{v_2} = 0.7$



Generally (hence also for constellations II and III, as shown in appendix 2), the optimal strategy for player 1 is to let v_1 approach $\widetilde{v_2}$. More specifically, if $\widetilde{v_1} > \widetilde{v_2}$, v_1 should remain larger than $\widetilde{v_2}$, and if $\widetilde{v_1} < \widetilde{v_2}$, v_1 should remain smaller than $\widetilde{v_2}$. This is the optimal manipulation strategy proposed by Schüssler (2007).

3.3 Optimal Manipulation and the Risk of Over-manipulation

In the case above, if player 1's optimal manipulation strategy works, her payoff against the honest player 2 is $P_1^{(b)max} = 1 - \frac{\widetilde{v_1}}{2\,\widetilde{v_2}}$. However, if she misjudges $\widetilde{v_2}$ and therefore manipulates 'too far', the payoff function switches to $P_1^{(a)}$, and her payoff for failed manipulation becomes $P_1^{fail} < \frac{\widetilde{v_1}}{2\widetilde{v_2}}$ since $P_1^{(a)}$ decreases with v_1. Therefore, $P_1^{fail} = \frac{\widetilde{v_1}}{2\widetilde{v_2}}$ is the most player 1 can expect from failed manipulation. More over-manipulation reduces her payoff further.

Compared to the strategy 'honesty' with $P_1^{honest} = 1 - \frac{\widetilde{v_1}}{\widetilde{v_1} + \widetilde{v_2}}$, player 1 could at the most gain $g = P_1^{(b)max} - P_1^{honest}$. On the other hand, if her manipulation fails, she will at least lose $l = P_1^{honest} - P_1^{fail}$. Calculating the difference between maximal potential gains and minimal potential losses results in the term $g - l = P_1^{max} + P_1^{fail} - 2\,P_1^{honest} = \frac{2\widetilde{v_1}}{\widetilde{v_1}+\widetilde{v_2}} - 1$. It is easy to see that $g - l = \frac{2\widetilde{v_1}}{\widetilde{v_1}+\widetilde{v_2}} - 1 < 0$, since in I, $\widetilde{v_1} < \widetilde{v_2}$.

This is a crucial result because it shows that the manipulation strategy which seeks to maximize player 1's gains through letting v_1 approach $\widetilde{v_2}$ has a higher potential loss through failure than can maximally be gained by successful manipulation. The same result is also found for constellations II and III, as shown in appendix 2.

This result can have a strong impact on a player's incentives to manipulate. If there is uncertainty about one's opponent's valuation, a player should be reluctant to follow the above optimal manipulation strategy. In real bargaining situations, at least some uncertainty can always be expected to be present. Manipulation therefore becomes less attractive.

Even if there is no uncertainty about valuations, a truthful player could protect herself by announcing a randomization strategy, which lets her announce her true valuation plus some error term. This could partly prevent the other player from manipulating, since she would increase the risk to lose more than she could gain. Such a strategy could be employed only if a player can credibly communicate such a strategy, or if players play a number of times.

To extend this type of reasoning, one would have to calculate the optimal strategy for settings of imperfect information. For example, one could assume a uniform (or normal) distribution of $\widetilde{v_2}$ over the interval [0;1] and check what player 1's optimal strategy would be for a given $\widetilde{v_1}$. However, this would go beyond the scope of this paper. The reasoning presented thus far shall nonetheless hint upon the fact that truth-telling might be a promising candidate for an optimal strategy in certain settings.

3.4 Mutual Manipulation

Previously, player 2 was assumed to announce her valuation truthfully. This section considers the situation of two manipulative players, which is in a way the worst case scenario when using AW. Yet, it is probably not only the most realistic assumption, but certainly the crucial test case for AW.

Schüssler (2007) considers the case for *n* objects and where players can choose between truth-telling and the optimal manipulation strategy from above. For this setting, he argues that the binary choice between manipulation and truth-telling renders the players in a chicken game. In this section I argue that this changes once manipulation is characterized in the continuous fashion proposed in this paper. Both players can choose not only whether or not to manipulate; they can also choose *how much* they want to manipulate, i.e. what valuation they want to announce.

The considerations from the previous section were already based on that assumption. The optimal strategy from above was only derived as the best response to a truthful player. The same result would also occur if one player had a first-mover's advantage: Through committing to the above optimal manipulation strategy, she could force the other player to state her true preferences, which is the best response to said strategy.

If the players have to announce their valuations simultaneously, the situation is very similar to Nash's demand game (Nash 1953). The announced valuation is the analog to the announced demand. This is an appropriate way of looking at AW as a strategic game because when valuations are supposed to be announced truthfully, open haggling should not be expected to occur. Mediators should further be able to enforce simultaneous valuation announcement, since one runs into much deeper troubles if this is not the case. From a theoretical point of view, this is also the simplest form of describing a symmetric situation in which neither player has a structural advantage. This is why this valuation-demand game shall be analyzed thereafter.

To begin with the strategic analysis, consider again constellation I from the previous section, where $\widetilde{v_2} > \widetilde{v_1} > 1 - \widetilde{v_2}$, depicted in Fig. 2. As long as $v_1 < v_2$, the players haggle about how much of object A is redistributed from player 2 to 1. These will be called the 'compatible' cases, because the initial distribution of objects remains efficient. As soon as $v_1 > v_2$, the manipulation strategies fail because payoff functions switch from (b) to (a) for player 1, and from (a) to (b) for 2. The outcome becomes inefficient.

Formalizing these intuitive arguments, a Nash-equilibrium in the valuation-demand game must fulfill the requirement that the two valuations are the same in the limit. If it is assumed that in case equal valuations are announced, the players decide who receives which object in AW step 2, they will always choose to allocate objects in accordance with the true valuations. This means that they will choose the better payoff function to apply for each of them. If they did not, both players would be worse off.

Hence every Nash equilibrium must fulfill the condition $v_1 = v_2 = v$. If, for instance in a situation of constellation I, player 1 would lower v_1 in comparison to the Nash equilibrium with $\widetilde{v_1} > v > \widetilde{v_2}$, she would render demands incompatible and decrease her payoff. By increasing v_1 back towards $\widetilde{v_1}$, she would also decrease her payoff since the amount of object A that she has to redistribute to 2 would become larger. Thus, no unilateral deviation from this strategy can be profitable for her, and a beneficial deviation is always possible if $v_1 = v_2 = v$ does not hold.

The main difference to this game and the original Nash demand game is the structure of payoffs. In the original demand game, payoffs are the respective demands themselves for compatible demands, and zero for both players in case of incompatibility. The payoffs here (for constellation I) are

- $P_1^{(b)} = \frac{\widetilde{v_1}}{v_1+v_2}$ and $P_2^{(a)} = 1 - \frac{\widetilde{v_2}}{v_1+v_2}$ for compatible demands with $v_1 \leq v_2$, and
- $P_1^{(a)} = 1 - \frac{\widetilde{v_1}}{v_1+v_2}$ and $P_2^{(b)} = \frac{\widetilde{v_2}}{v_1+v_2}$ for incompatible demands with $v_1 > v_2$.

The incompatibility payoff is not zero, it is not even constant in the AW-manipulation-game. Therefore, the question which Nash equilibrium will be selected cannot be answered unambiguously.

3.5 The Threat-equivalent Equilibrium

However, there is one Nash equilibrium with a special appeal to it. This is the equilibrium where both players face the same loss in case demands become incompatible. Call this the *threat-equivalent equilibrium*. In that equilibrium, both players have the same capacity to threaten the other player into behaving compatibly. In every other Nash equilibrium, one player faces a higher potential loss through incompatibility than the other player.

If any other equilibrium were to be chosen, one player could threaten the other and argue as following: 'If you do not reduce your demand, I will render mutual demands incompatible. You would lose more than me through my move, therefore I urge you to reduce your demand.' She could make this argument exactly up to the point where both could threaten each other with the same potential loss. This is the threat-equivalent equilibrium.

To calculate this equilibrium strategy, the equilibrium payoff of player 1 minus her potential loss at this equilibrium point must be equal to player 2's equilibrium payoff minus 2's potential loss.

$$P_1^{(a)}(v,v) - P_1^{(b)}(v,v) = P_2^{(b)}(v,v) - P_2^{(a)}(v,v)$$

$$\frac{\widetilde{v_1}}{2v} + \left(\frac{\widetilde{v_1}}{2v} - 1\right) = 1 - \frac{\widetilde{v_2}}{2v} - \frac{\widetilde{v_2}}{2v}$$

$$\frac{\widetilde{v_1}}{v} - 1 = 1 - \frac{\widetilde{v_2}}{v}$$

$$v = \frac{\widetilde{v_1} + \widetilde{v_2}}{2}$$

This solution has a simple graphic interpretation: The point where both face equal potential losses is exactly the point in the middle between the two true valuations: $\frac{\widetilde{v_1}+\widetilde{v_2}}{2}$. Both players shift their valuation the same distance from their true valuation towards the other's true valuation in the threat-equivalent equilibrium.

The very important feature of this equilibrium is that the payoffs for both players are exactly the same as in the case where both players play truthful strategies:

$$P_1^{(b)}\left(\frac{\widetilde{v_1} + \widetilde{v_2}}{2}, \frac{\widetilde{v_1} + \widetilde{v_2}}{2}\right) = 1 - \frac{\widetilde{v_1}}{2\left(\frac{\widetilde{v_1} + \widetilde{v_2}}{2}\right)} = 1 - \frac{\widetilde{v_1}}{\widetilde{v_1} + \widetilde{v_2}} = P_1^{(b)}(\widetilde{v_1}, \widetilde{v_2})$$

$$P_2^{(a)}\left(\frac{\widetilde{v_1} + \widetilde{v_2}}{2}, \frac{\widetilde{v_1} + \widetilde{v_2}}{2}\right) = \frac{\widetilde{v_2}}{2\left(\frac{\widetilde{v_1} + \widetilde{v_2}}{2}\right)} = \frac{\widetilde{v_2}}{\widetilde{v_1} + \widetilde{v_2}} = P_2^{(a)}(\widetilde{v_1}, \widetilde{v_2})$$

This result is truly remarkable. $P_2^{(a)}\left(\frac{\widetilde{v_1}+\widetilde{v_2}}{2}, \frac{\widetilde{v_1}+\widetilde{v_2}}{2}\right)$ is the payoffs under the prescribed threat equivalent manipulation. $P_2^{(a)}(\widetilde{v_1}, \widetilde{v_2})$ is the payoff if both are truthful. Those two payoffs are the same, which means that *if both players manipulate in this way, manipulation does not matter*, and the solution preserves all the appealing properties of the ideal AW procedure. In particular, it is efficient and fair under the definition of Kalai and Smorodinsky.

The justification that this equilibrium will be the outcome of the game is not without counter-arguments, and it is by no means the claim of this paper that the threat-equivalent equilibrium is the only feasible equilibrium outcome here. Nevertheless, the comparison of losses in that way is a reasonable argument to justify an equilibrium – especially since AW normalizes the maximal payoffs of both players, which means that they are comparable between the players. Thus, the argument of equivalent losses is a solid argument if players have similar risk preferences.

Furthermore, the threat-equivalent equilibrium has the appealing property that it is the symmetric point between the two true valuations. This could be another argument for a player to choose the according strategy, since both would be manipulating equally strong here.

3.6 Symmetric Manipulation

In the threat-equivalent equilibrium, both players manipulate equally strongly *and* they are in a Nash equilibrium. What if that latter characteristic is dropped, and it is only assumed that both players manipulate to the same extent?

To answer this question, first assume that both shift their true valuation *t* points towards the other's valuation, and demands remain compatible ($t < \frac{\widetilde{v_2}-\widetilde{v_1}}{2}$). As the calculation below shows, the payoffs as under truth telling are preserved. The equality $P_1^{(b)}(\widetilde{v_1}+t, \widetilde{v_2}-t) = P_1^{(b)}(\widetilde{v_1}, \widetilde{v_2})$ means that the payoffs where both manipulate with *t* are the same as when both announce truthfully. This is true for both players. Thus, *as long as demands remain compatible and manipulation is symmetric, manipulation does not matter either*.

$$P_1^{(b)}(\widetilde{v_1}+t, \widetilde{v_2}-t) = 1 - \frac{\widetilde{v_1}}{\widetilde{v_1}+t+\widetilde{v_2}-t} = 1 - \frac{\widetilde{v_1}}{\widetilde{v_1}+\widetilde{v_2}} = P_1^{(b)}(\widetilde{v_1}, \widetilde{v_2})$$

$$P_2^{(a)}(\widetilde{v_1}+t, \widetilde{v_2}-t) = \frac{\widetilde{v_2}}{\widetilde{v_1}+t+\widetilde{v_2}-t} = \frac{\widetilde{v_2}}{\widetilde{v_1}+\widetilde{v_2}} = P_2^{(a)}(\widetilde{v_1}, \widetilde{v_2})$$

Now consider the case where both manipulate equally strongly, but too strong to keep demands compatible ($t > \frac{\widetilde{v_2}-\widetilde{v_1}}{2}$). The obtained payoffs are as follows:

$$P_1^{(a)}(\widetilde{v_1}+t, \widetilde{v_2}-t) = \frac{\widetilde{v_1}}{\widetilde{v_1}+t+\widetilde{v_2}-t} = \frac{\widetilde{v_1}}{\widetilde{v_1}+\widetilde{v_2}}$$

$$P_2^{(b)}(\widetilde{v_1}+t, \widetilde{v_2}-t) = 1 - \frac{\widetilde{v_2}}{\widetilde{v_1}+t+\widetilde{v_2}-t} = 1 - \frac{\widetilde{v_2}}{\widetilde{v_1}+\widetilde{v_2}}$$

Those payoffs are certainly smaller than before. Comparing those payoffs with what they would have gotten under mutual truth telling (or in the threat-equivalent-equilibrium, or any case where they manipulate equally strong but remain compatible), one obtains:

$$P_1^{(a)}(\widetilde{v_1}+t,\widetilde{v_2}-t)-P_1^{(b)}(\widetilde{v_1},\widetilde{v_2})=\frac{\widetilde{v_1}}{\widetilde{v_1}+\widetilde{v_2}}-\left(1-\frac{\widetilde{v_1}}{\widetilde{v_1}+\widetilde{v_2}}\right)=\frac{2\,\widetilde{v_1}}{\widetilde{v_1}+\widetilde{v_2}}-1=\frac{\widetilde{v_1}-\widetilde{v_2}}{\widetilde{v_1}+\widetilde{v_2}}$$

$$P_2^{(b)}(\widetilde{v_1}+t,\widetilde{v_2}-t)-P_2^{(a)}(\widetilde{v_1},\widetilde{v_2})=1-\frac{\widetilde{v_2}}{\widetilde{v_1}+\widetilde{v_2}}-\left(\frac{\widetilde{v_2}}{\widetilde{v_1}+\widetilde{v_2}}\right)=1-\frac{2\,\widetilde{v_2}}{\widetilde{v_1}+\widetilde{v_2}}=\frac{\widetilde{v_1}-\widetilde{v_2}}{\widetilde{v_1}+\widetilde{v_2}}$$

Since $\widetilde{v_1}<\widetilde{v_2}$, both terms are negative, hence losses occur through over-manipulation (as before). Therefore, losses compared to honesty are the same for both players if both players over-manipulate to the same extent, namely $\frac{\widetilde{v_1}-\widetilde{v_2}}{\widetilde{v_1}+\widetilde{v_2}}$. The same results are also found for the remaining constellations, described in appendix 3.

In conclusion of this last result, if both players manipulate equally strongly, effects are symmetric. If they remain compatible, the payoffs from truthfulness and hence AW's appealing solution is preserved. If manipulation renders payoffs incompatible, both players face equal losses when they manipulate equally strongly.

To summarize the main results from section 3: In the threat-equivalent equilibrium, both players manipulate. Yet, the same payoffs as in the AW-solution and hence all its properties re preserved. Due to the strategic interaction, manipulation cancels out and the payoffs are exactly the same as under truth-telling. Further, even if players are not in equilibrium but manipulate symmetrically, the AW properties are still preserved as long as manipulation does not change the initial distribution of objects.

Thus, the calculations for this model show that there is quite a range of cases of unproblematic manipulation. There are good arguments why those situations are more likely to occur (symmetry, equal threats, equal losses). Manipulation has therefore less of a harmful impact on the AW mechanism.

4. Conclusion: Assessing the Problem of Manipulation for AW

The previous part has analyzed manipulation strategies for AW in the two objects case. Even though the problem of manipulation has not been shown to be negligible in all cases, there are circumstances where the problem is mitigated. *First*, since it has been shown that manipulation carries the danger of over-manipulation, imperfect information can be beneficial for the applicability of AW. The less knowledge a player has about an opponent's valuation, the riskier it is for her to manipulate. A player can create such uncertainty for her opponent by randomizing her own valuation, thereby creating the danger of over-manipulation for the opponent.

Second, if both players are expected to manipulate, there are still cases where manipulation has no negative effects on the outcome. As has been shown, there is an equilibrium which reproduces the payoffs from truthful AW. The same reasoning applies to all cases where both manipulate equally strongly as long as the initial allocation of objects does not change.

Thus, AW is not as severely threatened by manipulation as it seems at first sight. Imperfect information, randomization of one's own strategy and mutual strategic manipulation mitigate the problem. Still, further theoretical and practical research is needed to deepen the understanding of manipulative strategies in theory and practice. The approach taken here can serve as a starting point and a baseline approach for further modeling.

What do these findings imply for the AW method and its applicability? The main insight from these findings is that manipulation may not be as ubiquitous as one might expect at first sight. The reason for this is that manipulation performs badly in a cost-benefit-analysis. Whilst this point was already partly implied by Schüssler's model, a new finding in this paper is the fact that manipulation may not matter even though it occurs, and that also solutions which came about under manipulative behavior can still be fair. Thus, the AW mechanism can be seen as resilient against certain kinds untruthful behavior. This improves AW's applicability to dispute settlements in areas where agents must be assumed to be pure egoists, for instance in the realm of international politics. Admittedly, conflicts of a political importance comparable to the Israel-Egypt case above will probably not be settled by applying the AW procedure any time soon. Yet, even for those cases, AW may provide an interesting perspective on what a solution could look like. For other cases, such as the negotiation of coalition agreements, AW can provide a reasonable guideline, especially under time pressure, even when manipulative behavior most certainly occurs.

The next crucial theoretical step is the generalization of the approach to cases with more than two goods. This complicates matters more than one might imagine at first sight. It is possible that the additional complication of the situation favors truthful behavior as more uncertainty is introduced. Unfortunately, analytical results in that direction are hard to obtain, as exemplified by the paper by Aziz et al (2015): The authors provide some general insights for the n-goods case, most prominently they prove the existence of Nash-equilibria under certain conditions. Yet, they are only able to characterize the impact on general welfare of those equilibria, which they show to be at least ¾ of the welfare from the original solution. They do not identify characteristics of those solutions with regards to how this welfare is split. Thus, the question about fairness in AW with manipulation in the n-goods case remains unanswered.

A solution to those difficulties might be found in the computer simulation of AW. For example, an agent-based model in which manipulating and truthful players compete in multiple issue negotiations could be employed to assess which strategies prove to be most successful.

Appendix 1: Calculation of the Payoff Functions of AW

This section gives the calculations for all cases that can occur for the payoff function of a player in AW with two objects. These results are referred to in section 0.

Case (a)

Relevant constraints: $v_1 > v_2$ and $v_1 > 1 - v_2$.

Initial allocation of objects: Player 1 receives object A; Player 2 receives object B.

Redistribution: Since $v_1 > 1 - v_2$, redistribution of x percent of A from 1 to 2.

Calculation of x: $$(1 - x) * v_1 = (1 - v_2) + x * v_2$$

$$x = \frac{v_1 + v_2 - 1}{v_1 + v_2} = 1 - \frac{1}{v_1 + v_2}$$

Note that for the calculation of x, the announced valuations are relevant. Below, the calculation of the payoff also depends on the true valuation.

Payoff for player 1: $$P_1^{(a)} = (1 - x) * \widetilde{v_1}$$

$$P_1^{(a)} = \left(1 - (1 - \frac{1}{v_1 + v_2})\right) * \widetilde{v_1}$$

$$P_1^{(a)} = \frac{\widetilde{v_1}}{v_1 + v_2}$$

Case (b)

Relevant constraints: $v_1 < v_2$ and $1 - v_1 < v_2$.

Initial allocation of objects: Player 1 receives object B; Player 2 receives object A.

Redistribution: Since $1 - v_1 < v_2$, redistribution of x percent of A from 2 to 1.

Calculation of x: $$(1 - v_1) + x * v_1 = (1 - x) * v_2$$

$$x = \frac{v_1 + v_2 - 1}{v_1 + v_2} = 1 - \frac{1}{v_1 + v_2}$$

Payoff for player 1: $$P_1^{(b)} = (1 - \widetilde{v_1}) + x * \widetilde{v_1}$$

$$P_1^{(b)} = (1 - \widetilde{v_1}) + (1 - \frac{1}{v_1 + v_2}) * \widetilde{v_1}$$

$$P_1^{(b)} = 1 - \frac{\widetilde{v_1}}{v_1 + v_2}$$

Case (c)

Relevant constraints: $v_1 > v_2$ and $v_1 < 1 - v_2$.

Initial allocation of objects: Player 1 receives object A; Player 2 receives object B.

Redistribution: Since $v_1 < 1 - v_2$, redistribution of x percent of B from 2 to 1.

Calculation of x: $v_1 + x * (1 - v_1) = (1 - x) * (1 - v_2)$

$$x = \frac{1-v_1-v_2}{2-v_1-v_2}$$

Payoff for player 1: $P_1^{(c)} = \widetilde{v_1} + x * (1 - \widetilde{v_1})$

$$P_1^{(c)} = \widetilde{v_1} + \frac{1-v_1-v_2}{2-v_1-v_2} * (1 - \widetilde{v_1})$$

$$P_1^{(c)} = \widetilde{v_1} * (1 - \frac{1-v_1-v_2}{2-v_1-v_2}) + \frac{1-v_1-v_2}{2-v_1-v_2}$$

$$P_1^{(c)} = \widetilde{v_1}(\frac{1}{2-v_1-v_2}) + \frac{1-v_1-v_2}{2-v_1-v_2}$$

$$P_1^{(c)} = \frac{\widetilde{v_1}+1-v_1-v_2}{2-v_1-v_2} = \frac{\widetilde{v_1}-1}{2-v_1-v_2} + 1 = 1 - \frac{1-\widetilde{v_1}}{(1-v_1)+(1-v_2)}$$

Case (d)

Relevant constraints: $v_1 < v_2$ and $1 - v_1 > v_2$.

Initial allocation of objects: Player 1 receives object B; Player 2 receives object A.

Redistribution: Since $1 - v_1 > v_2$, redistribution of x percent of B from 1 to 2.

Calculation of x: $(1 - x) * (1 - v_1) = v_2 + x * (1 - v_2)$

$$1 - x - v_1 + x * v_1 = v_2 + x - x * v_2$$

$$1 - v_1 - v_2 = 2x - x * v_1 - x * v_2$$

$$x = \frac{1-v_1-v_2}{2-v_1-v_2}$$

Payoff for player 1: $P_1^{(c)} = (1 - x) * (1 - \widetilde{v_1})$

$$P_1^{(c)} = \left(1 - \frac{1-v_1-v_2}{2-v_1-v_2}\right) * (1 - \widetilde{v_1})$$

$$P_1^{(c)} = \frac{1-\widetilde{v_1}}{2-v_1-v_2} = \frac{1-\widetilde{v_1}}{(1-v_1)+(1-v_2)}$$

Appendix 2: Optimal manipulation, potential gains and losses

This section completes the calculations for the result that the optimal strategy for player 1 against an honest player 2 is to let v_1 approach $\widetilde{v_2}$ for constellations II and III. Also, the result that the potential gains from optimal manipulation are always smaller than the potential losses from over-manipulation are shown for constellations II and III.

Optimal strategy in Constellation II

$\widetilde{v_2}$ lies in II, hence $1 - \widetilde{v_1} < \widetilde{v_2} < \widetilde{v_1}$. Depending on 1's choice of v_1, either payoff function (a), (b), or (d) is applicable. The payoff functions and the local maxima are the same as in constellation I, but the global maximum is different:

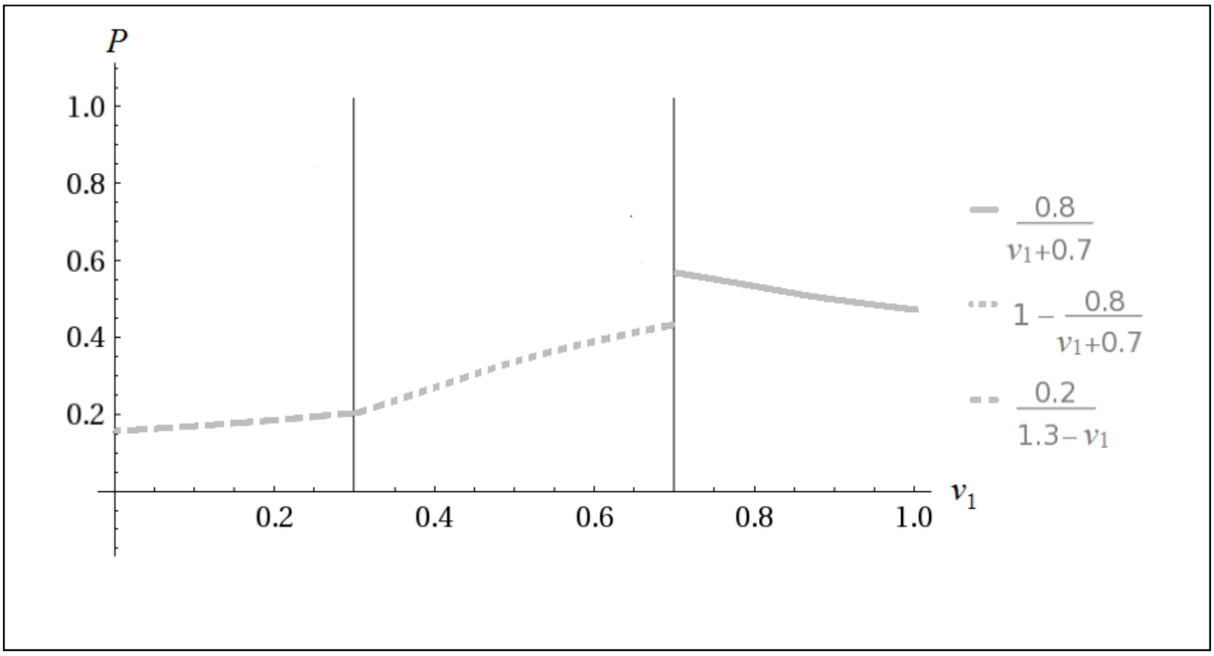
- Local maximum in (a): $P_1^{(a)max} = \frac{\widetilde{v_1}}{2\,\widetilde{v_2}}$ for $v_1 = \widetilde{v_2}$
- Local maximum in (b): $P_1^{(b)max} = 1 - \frac{\widetilde{v_1}}{2\,\widetilde{v_2}}$ for $v_1 = \widetilde{v_2}$
- Local maximum in (d): $P_1^{(d)max} = (1 - \widetilde{v_1})$ for $v_1 = 1 - \widetilde{v_2}$

From $1 - \widetilde{v_1} < \widetilde{v_2} < \widetilde{v_1}$ follows

$$P_1^{(a)max} > P_1^{(b)max} \text{ and } P_1^{(a)max} > P_1^{(d)max}.$$

The global maximum is $P_1^{max} = P_1^{(a)max}$. Therefore, player 1's optimal strategy is to choose $v_1 = \widetilde{v_2}$ but still a little above $\widetilde{v_2}$ to stay in (a) (technically correct: $v_1 = \widetilde{v_2} + \varepsilon$ with $\varepsilon \to 0$ and $\varepsilon > 0$). Fig. 3 illustrates the payoff function for the example $\widetilde{v_1} = 0.8$ and $\widetilde{v_2} = 0.7$.

Fig. 3: Exemplary payoff function for constellation II with $\widetilde{v_1} = 0.8$ and $\widetilde{v_2} = 0.7$



Gains and losses from manipulation in constellation II

If player 1's optimal manipulation strategy works, her payoff against the honest player 2 is

$$P_1^{(a)max} = \frac{\widetilde{v_1}}{2\,\widetilde{v_2}}.$$

If he only slightly manipulates too much, the payoff function switches to $P_1^{(b)}$, and her maximal payoff for failed manipulation is

$$P_1^{(b)fail} = 1 - \frac{\widetilde{v_1}}{2\widetilde{v_2}}.$$

The payoff from strategy 'honesty' is

$$P_1^{(a)honest} = \frac{\widetilde{v_1}}{\widetilde{v_1} + \widetilde{v_2}}.$$

Player 1 could at the most gain $g = P_1^{(a)max} - P_1^{(a)honest}$. On the other hand, if her manipulation fails, he will at least lose $l = P_1^{(a)honest} - P_1^{(b)fail}$. Calculating the difference between maximal potential gains and minimal potential losses results in the term

$$g - l = P_1^{(a)max} + P_1^{(b)fail} - 2\,P_1^{(a)honest} = 1 - \frac{2\widetilde{v_1}}{\widetilde{v_1}+\widetilde{v_2}}.$$

In II, $\widetilde{v_1} > \widetilde{v_2}$, and therefore

$$g - l = 1 - \frac{2\widetilde{v_1}}{\widetilde{v_1}+\widetilde{v_2}} < 0.$$

Optimal strategy in constellation III:

$\widetilde{v_2}$ lies in III, hence $\widetilde{v_2} < 1 - \widetilde{v_1} < \widetilde{v_1} < 1 - \widetilde{v_2}$. Depending on 1's choice of v_1, either payoff function (a), (c), or (d) is applicable. The payoff functions are *different* than in constellation I and II. Also, local and global maxima are different than in I and II.

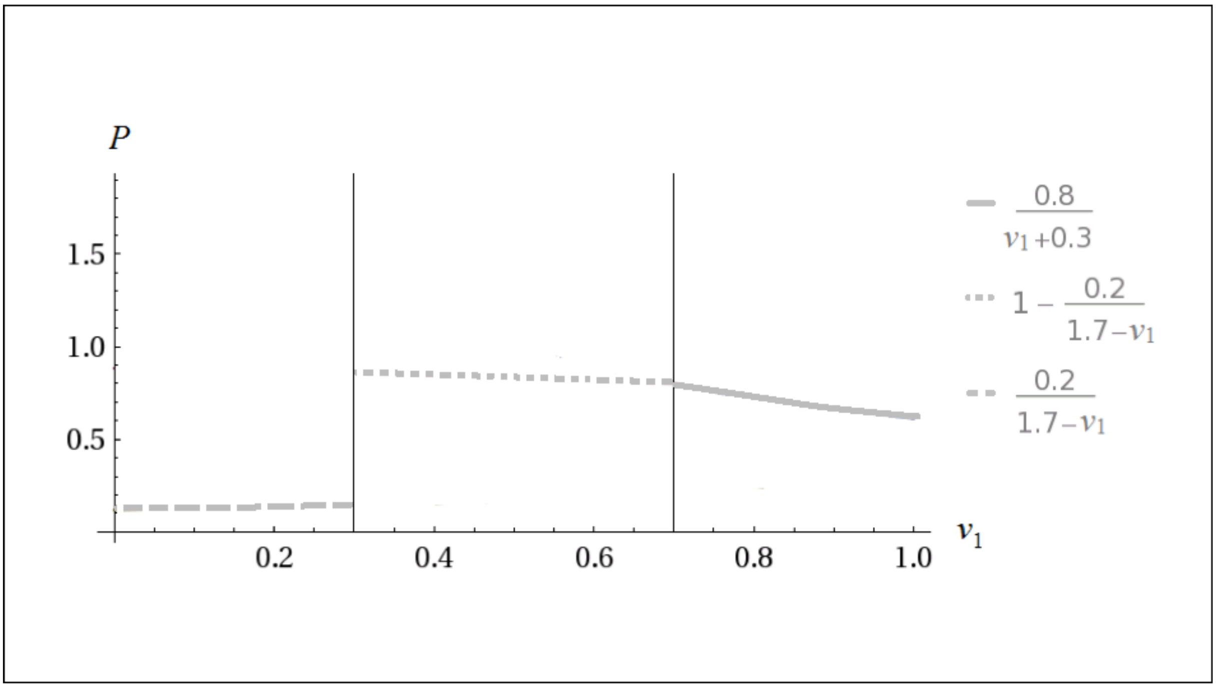
- For $v_1 > 1 - \widetilde{v_2}$: $P_1^{(a)} = \frac{\widetilde{v_1}}{v_1 + \widetilde{v_2}}$; which is strictly decreasing with v_1; local maximum in (a): $P_1^{(a)max} = \widetilde{v_1}$ for $v_1 = 1 - \widetilde{v_2}$
- For $\widetilde{v_2} < v_1 < 1 - \widetilde{v_2}$: $P_1^{(c)} = 1 - \frac{(1-\widetilde{v_1})}{(1-v_1)+(1-\widetilde{v_2})}$; which is strictly decreasing with v_1; local maximum in (c): $P_1^{(c)max} = 1 - \frac{(1-\widetilde{v_1})}{2(1-\widetilde{v_2})}$ for $v_1 = \widetilde{v_2}$
- For $v_1 < \widetilde{v_2}$: $P_1^{(d)} = \frac{1-\widetilde{v_1}}{(1-v_1)+(1-\widetilde{v_2})}$, which is strictly increasing with v_1; local maximum in (d): $P_1^{(d)max} = \frac{(1-\widetilde{v_1})}{2(1-\widetilde{v_2})}$ for $v_1 = \widetilde{v_2}$

From $\widetilde{v_2} < 1 - \widetilde{v_1} < \widetilde{v_1} < 1 - \widetilde{v_2}$ follows

$$P_1^{(c)max} > P_1^{(d)max} \text{ and } P_1^{(c)max} > P_1^{(a)max}.$$

To see the latter, calculate $P_1^{(c)max} - P_1^{(a)max} = \frac{(1-\widetilde{v_1})*(1-2\widetilde{v_2})}{2(1-\widetilde{v_2})} > 0$ (since $\widetilde{v_2} < 0.5$, all factors are < 0. The global maximum is therefore $P_1^{max} = P_1^{(c)max}$. This means that player 1's optimal strategy is $v_1 = \widetilde{v_2}$ but still a little above $\widetilde{v_2}$ to stay in (c) (technically correct: $v_1 = \widetilde{v_2} + \varepsilon$ with $\varepsilon \to 0$ and $\varepsilon > 0$). Fig. 4 illustrates the payoff function for $\widetilde{v_1} = 0.8$ and $\widetilde{v_2} = 0.3$.

Fig. 4: Exemplary payoff function for constellation III with $\widetilde{v_1} = 0.8$ and $\widetilde{v_2} = 0.3$



Gains and losses from manipulation in constellation III

If player 1's optimal manipulation strategy works, her payoff against the honest player 2 is

$$P_1^{(c)max} = 1 - \frac{(1-\widetilde{v_1})}{2(1-\widetilde{v_2})}.$$

If he only slightly manipulates too much, the payoff function switches to $P_1^{(d)}$, and her maximal payoff for failed manipulation is

$$P_1^{(d)fail} = \frac{(1-\widetilde{v_1})}{2(1-\widetilde{v_2})}.$$

The payoff from strategy 'honesty' is

$$P_1^{(c)honest} = 1 - \frac{(1-\widetilde{v_1})}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}.$$

Player 1 could at the most gain $g = P_1^{max} - P_1^{honest}$. On the other hand, if her manipulation fails, he will at least lose $l = P_1^{honest} - P_1^{fail}$. Calculating the difference between maximal potential gains and minimal potential losses results in the term

$$g - l = P_1^{(c)max} + P_1^{fail} - 2\,P_1^{honest} = \frac{2(1-\widetilde{v_1})}{(1-\widetilde{v_1})+(1-\widetilde{v_2})} - 1.$$

In III, $1-\widetilde{v_2} > 1-\widetilde{v_1}$, and therefore

$$g - l = \frac{2(1-\widetilde{v_1})}{(1-\widetilde{v_1})+(1-\widetilde{v_2})} - 1 < 0.$$

Appendix 3: Calculations for the threat-equivalent equilibrium for other constellations

Constellation III

Consider first true valuation constellation III with $\widetilde{v_1} > \widetilde{v_2}$ and $1-\widetilde{v_2} > \widetilde{v_1}$ (See Fig. 2). Again, both players manipulate their valuation towards the other's true valuation, and also for every Nash equilibrium $v_1 = v_2 = v$. However, two types of equilibria need to be distinguished in III, namely either $v > 0.5$ or $v < 0.5$.

In the case $v > 0.5$, the payoff functions are the same as in constellation I, only with reversed roles of the players. The proof for the threat-equivalent equilibrium and the resulting payoff therefore remains the same.

In the case $v < 0.5$, the payoff functions are

- $P_1^{(c)} = 1 - \frac{1-\widetilde{v_1}}{(1-v_1)+(1-v_2)}$ and $P_2^{(d)} = \frac{1-\widetilde{v_2}}{(1-v_1)+(1-v_2)}$ for compatible demands $v_1 > v_2$
- $P_1^{(d)} = \frac{1-\widetilde{v_1}}{(1-v_1)+(1-v_2)}$ and $P_2^{(c)} = 1 - \frac{1-\widetilde{v_2}}{(1-v_1)+(1-v_2)}$ for incompatible demands $v_1 < v_2$.

The threat-equivalent equilibrium is then calculated along similar lines:

$$P_1^{(c)}(v,v) - P_1^{(d)}(v,v) = P_2^{(d)}(v,v) - P_2^{(c)}(v,v)$$

$$1 - \frac{1-\widetilde{v_1}}{2(1-v)} - \frac{1-\widetilde{v_1}}{2(1-v)} = \frac{1-\widetilde{v_2}}{2(1-v)} - (1 - \frac{1-\widetilde{v_2}}{2(1-v)})$$

$$1 - \frac{1-\widetilde{v_1}}{1-v} = \frac{1-\widetilde{v_2}}{(1-v)} - 1$$

$$v = \frac{\widetilde{v_1} + \widetilde{v_2}}{2}$$

The payoffs are again the same as from truth-telling.

$$P_1^{(c)}\left(\frac{\widetilde{v_1}+\widetilde{v_2}}{2}, \frac{\widetilde{v_1}+\widetilde{v_2}}{2}\right) = 1 - \frac{1-\widetilde{v_1}}{2-2\left(\frac{\widetilde{v_1}+\widetilde{v_2}}{2}\right)} = 1 - \frac{1-\widetilde{v_1}}{(1-\widetilde{v_1})+(1-\widetilde{v_2})} = P_1^{(c)}(\widetilde{v_1}, \widetilde{v_2})$$

$$P_2^{(d)}\left(\frac{\widetilde{v_1}+\widetilde{v_2}}{2}\right) = \frac{1-\widetilde{v_2}}{2-2\left(\frac{\widetilde{v_1}+\widetilde{v_2}}{2}\right)} = \frac{1-\widetilde{v_2}}{(1-\widetilde{v_1})+(1-\widetilde{v_2})} = P_2^{(d)}(\widetilde{v_1}, \widetilde{v_2})$$

Now consider symmetric manipulation. For compatible manipulation ($t < \frac{\widetilde{v_2}-\widetilde{v_1}}{2}$) in constellation III:

$$P_1^{(c)}(\widetilde{v_1}+t, \widetilde{v_2}-t) = 1 - \frac{(1-\widetilde{v_1})}{\big(1-(\widetilde{v_1}+t)\big)+\big(1-(\widetilde{v_2}-t)\big)} = 1 - \frac{(1-\widetilde{v_1})}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}$$
$$= P_1^{(c)}(\widetilde{v_1}, \widetilde{v_2})$$

$$P_2^{(d)}(\widetilde{v_1}+t, \widetilde{v_2}-t) = \frac{1-\widetilde{v_2}}{(1-(\widetilde{v_1}+t))+(1-(\widetilde{v_2}-t))} = \frac{1-\widetilde{v_2}}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}$$
$$= P_2^{(d)}(\widetilde{v_1}, \widetilde{v_2})$$

For incompatible manipulation ($t > \frac{\widetilde{v_2}-\widetilde{v_1}}{2}$) in constellation III, payoff functions are once again reversed.

$$P_1^{(d)}(\widetilde{v_1}+t,\widetilde{v_2}-t)=\frac{1-\widetilde{v_1}}{(1-(\widetilde{v_1+t}))+(1-(\widetilde{v_2}-t))}$$

$$P_2^{(c)}(\widetilde{v_1}+t,\widetilde{v_2}-t)=1-\frac{(1-\widetilde{v_2})}{\left(1-(\widetilde{v_1}+t)\right)+\left(1-(\widetilde{v_2}-t)\right)}=1-\frac{(1-\widetilde{v_2})}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}$$

Comparing this with the payoffs from mutual truthfulness

$$P_1^{(d)}(\widetilde{v_1}+t,\widetilde{v_2}-t)-P_1^{(c)}(\widetilde{v_1},\widetilde{v_2})=\frac{1-\widetilde{v_1}}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}-\left(1-\frac{1-\widetilde{v_1}}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}\right)$$
$$=\frac{2*(1-\widetilde{v_1})}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}-1=\frac{2*(1-\widetilde{v_1})-((1-\widetilde{v_1})+(1-\widetilde{v_2}))}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}$$
$$=\frac{(1-\widetilde{v_1})-(1-\widetilde{v_2})}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}$$

$$P_2^{(c)}(\widetilde{v_1}+t,\widetilde{v_2}-t)-P_2^{(d)}(\widetilde{v_1},\widetilde{v_2})=1-\frac{1-\widetilde{v_2}}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}-\left(\frac{1-\widetilde{v_2}}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}\right)$$
$$=1-\frac{2*(1-\widetilde{v_2})}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}=\frac{(1-\widetilde{v_1})+(1-\widetilde{v_2})-2*(1-\widetilde{v_2})}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}$$
$$=\frac{(1-\widetilde{v_1})-(1-\widetilde{v_2})}{(1-\widetilde{v_1})+(1-\widetilde{v_2})}$$

The potential losses are hence also shown to be of the same magnitude from symmetric over-manipulation.

Constellation II

What is now left to complete the result for all cases is constellation II with $\widetilde{v_1}>\widetilde{v_2}$ and $1-\widetilde{v_2}<\widetilde{v_1}$. No new calculations are necessary: If, in equilibrium, $v>0.5$, payoff functions are the same as in constellation I, but with reversed roles of the players. If $v<0.5$, the payoff functions are the same as in the case $v<0.5$ in constellation III. If not in equilibrium, due to the assumed symmetry of manipulation, no switches in payoff functions occur and hence the same logic can be applied also to those cases. The result is now proven for all possible constellations.

Literature

Aziz, Haris, Simina Brânzei, Aris Filos-Ratsikas, and Søren Kristoffer Stiil Frederiksen. 2015. *The Adjusted Winner Procedure: Characterizations and Equilibria, arxiv.org [arXiv:1503.06665].*

Brams, Steven and Alan Taylor. 1999: *The win-win solution. Guaranteeing fair shares to everybody.* New York: W.W. Norton (1st ed.).

Kalai, Ehud and Meir Smorodinsky. 1975. Other solutions to Nash's bargaining problem, *Econometrica* 43: 513–518.

Luce, R. and Howard Raiffa. 1989. *Games and decisions. Introduction and critical survey*. New York: Dover Publishers.

Nash, John. 1950. The bargaining problem, *Econometrica* 18: 155–162.

Nash, John. 1953: Two-Person Cooperative Games, *Econometrica* 21 (1): 128–140.

Raith, Matthias. 2000. Fair-negotiation procedures, *Mathematical Social Sciences* 39 (3): 303–322.

Rawls, John. 1985. Justice as Fairness: Political not Metaphysical. In: *Philosophy and Public Affairs 14*: 223-251.

Schüssler, Rudolf. 2007. "Adjusted Winner" (AW) Analyses of the 1978 Camp David Accords – Valuable Tools for Negotiators? In *Rudolf Avenhaus and I. William Zartman (Hg.): Diplomacy games. Formal models and international negotiations: 284 – 296.* Berlin: Springer.