

TCER Working Paper Series

A Methodological Note on a Weighted Voting Experiment

Eric Guerci  
Nobuyuki Hanaki  
Naoki Watanabe  
Gabriele Esposito  
Xiaoyan Lu

December 2013

Working Paper E-63

<http://tcer.or.jp/wp/pdf/e63.pdf>



TOKYO CENTER FOR ECONOMIC RESEARCH  
1-7-10-703 Iidabashi, Chiyoda-ku, Tokyo 102-0072, Japan

## Abstract

We conducted a sensitivity analysis of the results of weighted voting experiments by varying two features of the experimental protocol by Montero et al. (2008): (1) the way in which the roles of subjects are reassigned in each round (random role, RR, versus fixed role, FR) and (2) the number of proposals that subjects can simultaneously approve (multiple approval, MA, versus single approval, SA). It was observed that the differences in these protocols had impacts on the relative frequencies of minimum winning coalitions as well as how negotiations proceed. Our analysis favors a protocol with FR-SA for future research, because this protocol induces subjects to commit fewer errors in their decision making than the protocol with RR-MA, and because proposal-objection dynamics are more frequently observed under FR-SA than under RR-MA.

Eric Guerci  
Universite de Nice Sophia Antipolis  
Institut Supérieur D'Economie & Management  
Campus de St Jean d'Angely 24, av. des  
Diables Bleus 06357 NICE Cedex 04, France  
eric.guerci@gredeg.cnrs.fr

Nobuyuki Hanaki  
Aix-Marseille University  
School of Economics  
3, av. Robert Schuman 13628 Aix-en-Provence  
Cedex 1, France  
nobuyuki.HANAKI@univ-amu.fr

Naoki Watanabe  
University of Tsukuba  
faculty of Engineering, Information, and Systems  
1-1-1 Tennodai, Tsukuba, 305-8573, Japan  
naoki50@sk.tsukuba.ac.jp

Gabriele Esposito  
European Commission  
Joint Research Centre, Institute for  
Prospective Technological Studies  
Inca Garcilaso, s/n 41092 Seville, Spain  
gabriele.esposito@ec.europa.eu

Xiaoyan Lu  
Johnson & Johnson  
Janssen Pharmaceutica NV  
Turnhoustseweg 30, B-2340 Beerse, Belgium  
xlu2@its.jnj.com

# A methodological note on a weighted voting experiment\*

Eric Guerci<sup>†</sup>   Nobuyuki Hanaki<sup>‡</sup>   Naoki Watanabe<sup>§</sup>   Gabriele Esposito<sup>¶</sup>  
Xiaoyan Lu<sup>||</sup>

December 2, 2013

## Abstract

We conducted a sensitivity analysis of the results of weighted voting experiments by varying two features of the experimental protocol by Montero et al. (2008): (1) the way in which the roles of subjects are reassigned in each round (random role, RR, versus fixed role, FR) and (2) the number of proposals that subjects can simultaneously approve (multiple approval, MA, versus single approval, SA). It was observed that the differences in these protocols had impacts on the relative frequencies of minimum winning coalitions as well as how negotiations proceed. Our analysis favors a protocol with FR-SA for future research, because this protocol induces subjects to commit fewer errors in their decision making than the protocol with RR-MA, and because proposal-objection dynamics are more frequently observed under FR-SA than under RR-MA.

*JEL Classification Numbers:* C71, C92, D72

*Keywords:* weighted voting, experiment, cooperative game, winning coalition

---

\*This paper is a revised version of our earlier work “A Note on the Experiments of Weighted Voting: Human Mistakes in Cooperative Games” presented at TCER conference 2011. We have greatly benefited from comments and suggestions from Maria Montero. A part of this research was conducted while Hanaki was visiting the Higher School of Economics (Moscow). We have benefited from discussion with Fuad Aleskerov and Alexis Beliani. The Experimental Economics Lab at Montpellier (LAMETA), in particular, Dimitri Dubois, has kindly hosted us for running this experiment. Financial support from IEF Marie Curie research fellowship n.237633-MMI (Guerci), Japan Economic Research Foundation (Hanaki), MEXT Grants-in-Aid 20330037 and 25380222 (Watanabe), and JSPS-CNRS bilateral research project, JSPS-ANR bilateral research project “BECO” are gratefully acknowledged. *The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.*

<sup>†</sup>Université de Nice Sophia Antipolis (I.S.E.M.), CNRS GREDEG (UMR 7321). E-mail: eric.guerci@gredege.cnrs.fr

<sup>‡</sup>Aix-Marseille University (Aix Marseille School of Economics), CNRS, EHESS, and I.U.F. E-mail: nobuyuki.HANAKI@univ-amu.fr

<sup>§</sup>Corresponding author. Faculty of Engineering, Information, and Systems. University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan. Phone & Fax: +81-29-853-5166, E-mail: naoki50@sk.tsukuba.ac.jp.

<sup>¶</sup>European Commission, Joint Research Centre, Institute for Prospective Technological Studies. E-mail: gabriele.esposito@ec.europa.eu.

<sup>||</sup>Janssen Pharmaceutica NV. E-mail: xlu2@its.jnj.com.

# 1 Introduction

Weighted voting, which gives different numbers of votes (voting weights) to different voters, is a popular collective decision-making system in many contexts, including stockholder voting in corporations and multi-party legislatures. In such a system, however, the distribution of voters' actual voting powers is often remarkably different from the distribution of their nominal voting weights.<sup>1</sup> The relationship between the distribution of voting weights and voting powers is too complex to be foreseen, although several theoretical indices that measure "a priori" powers of voters, e.g., the Shapley-Shubik (SS) index (Shapley and Shubik, 1954), Banzhaf index (Banzhaf, 1965), and Deegan-Packel (DP) index (Deegan and Packel, 1978), have been proposed.

For a better understanding of this complex relationship inherent in weighted voting, the experimental study is currently attracting a good deal of researcher interest as a complement to empirical analysis, because many features that are unobservable in actual practice can be controlled in experiments. Montero et al. (2008) and Aleskerov et al. (2009) conducted the experiments in a cooperative game environment in which no explicit structure as to how negotiations take place is specified for voters' decision on how to allocate a fixed amount of resources via weighted voting.<sup>2</sup> An empirical measure of "a posteriori" power of a voter was defined as the average payoff that the voter obtained in certain rounds of the experimental session. In their experiments, it was observed that subjects eventually learned to form minimum winning coalitions (MWCs), i.e., winning coalitions such that a deviation of any member of the coalition turns the coalition status from winning to losing.<sup>3</sup>

The empirical measure of voting powers presents, however, dramatically different outcomes from the prediction that the theoretical indices (SS index, Banzhaf index, and DP index) suggest. In a 4-player weighted voting game  $[5; 1, 2, 2, 3]$  for allocating a total of 120 points, where the quota is 5 and the list of voting weights of the four voters is  $(1, 2, 2, 3)$ , the SS and Banzhaf indices suggest voting powers of 10, 30, 30, and 50 to the players with 1, 2, 2, and 3 votes, respectively, and the DP index suggests voting powers of 13.3, 33.3, 33.3, 40. In contrast, the empirical measure computed with data in the last 4 (out of 10) rounds of an experiment indicates 0.9, 25.1, 25.1, and 68.9 for the players with 1, 2, 2, and 3 votes, respectively (Montero et al., 2008, Table 2).

---

<sup>1</sup>Felsenthal and Machover (1998), for example, noted this point in their study of the Council of Ministers in the European Economic Community.

<sup>2</sup>Fréchette, et al. (2005b), Drouvelis et al. (2010), and Kagel et al. (2010) conducted their experiments in non-cooperative game environments based on variants of a legislative bargaining model by Baron and Ferejohn (1989). The analysis by Fréchette et al. (2005a) was based on the demand bargaining model by Morelli (1999).

<sup>3</sup>The formation of MWCs is consistent with the "size principle" proposed by Riker (1962), but the main purpose of the experiment conducted by Montero et al. (2008) was to test the "paradox of new members" i.e., adding a new member to a committee can increase the voting powers of some existing members, while reducing the powers of others. Aleskerov et al. (2009) tested the effect of asymmetry regarding favorable coalition partners among committee members on their coalition formation and resource allocation.

This remarkable discrepancy arises from the fact that a MWC of three voters (one player with 1 vote and two players with 2 votes each) was observed only once, while there were 58 observations for two MWCs of two voters (one player with 3 votes and one of the two players with 2 votes) among 64 observations. Namely, the observed relative frequencies of winning coalitions were far from those presumed by the theoretical indices; among three theoretical indices noted above, only the DP index is computed with a focus on MWCs, but it presumes that all the possible MWCs are equally likely to be formed and that resources are allocated equally among the members within each MWC.<sup>4</sup> Then, are there any regularities in experimental observations regarding (i) the relative frequencies with which possible MWCs are formed and (ii) how resources are allocated among the members within each MWC?

This note is the first step toward filling the gap between the empirical measure and the theoretical indices of voting powers by accumulating data that can give answers to questions (i) and (ii). Montero et al. (2008) and Aleskerov et al. (2009) required in their experimental protocol random reassignment of subjects' roles (RR) in each round of sessions and multiple approval (MA) of proposals that subjects can make simultaneously at a time. With small changes in protocol, however, Esposito et al. (2012) observed that MWCs of three voters appeared more frequently than those in earlier research. Although they examined different 4-player games, these games had three possible MWCs (one of three voters and two of two voters) as in the precedent research. The protocol of Esposito et al. (2012) required that the roles of subjects be partially fixed (semi-FR) and that each subject be in favor of only one proposal including his or her own at a time, i.e., single approval (SA).<sup>5</sup> It is thus of a considerable value to conduct a sensitivity analysis to identify the factors that generate different results while keeping the game intact, before starting to search for answers to questions (i) and (ii).

In our experiment two approval schemes, MA and SA, are considered as well as two role reassignments, RR and FR. In FA the roles of subjects are fixed throughout a session. We examined two weighted voting games with four voters that were used in Esposito et al. (2012). It was observed that the differences in those protocols had impacts on the relative frequencies of MWCs as well as how negotiations proceed. Our analysis of subject behavior favors a protocol with FR-SA for the accumulation of data to be used for future research, because this protocol induces subjects to commit fewer errors in their decision making than the protocol with RR-MA. FR-SA is also preferred to RR-MA for research on weighted voting in a cooperative game environment, because proposal-objection dynamics are more frequently observed under FR-SA than under RR-MA.

---

<sup>4</sup>Packel and Deegan (1980) generalized the DP index by giving a system of axioms for the weighted DP index.

<sup>5</sup>In Esposito et al. (2012), the number of votes of some, but not all, subjects varied when subjects were faced with the same game, because their focus was on learning about the relationship between voting weights and voting powers.

Table 1: The four treatments (protocols) examined.

	Random Role	Fixed Role
Multiple Approval	RR-MA	FR-MA
Single Approval	RR-SA	FR-SA

The rest of the paper is organized as follows. Section 2 describes the experimental design and introduces the above four protocols as treatments. Section 3 summarizes the experimental results with statistical analyses. Section 4 concludes this paper and also discusses other observations; in the game  $[5; 1, 2, 2, 3]$  examined in earlier research, MWCs of two voters were disproportionately more likely to be observed than those of three voters even under FR-SA.

## 2 Experimental design

Let  $N = \{1, 2, 3, 4\}$  be the set of players (voters). The players decide to allocate a fixed amount of resources via weighted voting. A 4-player weighted voting game is represented by  $[q; v_1, v_2, v_3, v_4]$ , where  $q$  is the quota (the least number of votes required to implement a proposal) and  $v_i$  is the voting weight (the number of votes) player  $i \in N$  has. A nonempty subset  $S$  of  $N$  is called a coalition, and it is called a winning coalition if  $\sum_{i \in S} v_i \geq q$ ; otherwise, it is called a losing coalition. A minimum winning coalition (MWC) is a winning coalition (WC) such that a deviation of any member of the coalition turns its status from winning to losing. For the experiment, we chose two weighted voting games previously examined in Esposito et al. (2012); (a)  $[14; 4, 4, 6, 8]$  and (b)  $[14; 7, 7, 3, 5]$ . (In both games, there is no veto player, i.e., the core is empty in each game.<sup>6</sup>)

Each session consists of 20 rounds. One of games (a) and (b) is played in the first 10 rounds and the other game is played in the second 10 rounds. In each round, subjects either propose how to allocate 100 points among the four members, approve others' proposals, or keep silent. When a subject approves a proposal, all of the subject's votes are cast for the proposal. A proposal is automatically approved by the proposer. The proposal that first obtains as many votes as the quota or more is adopted as the allocation of points among the members, and then the round is terminated. The protocol used by Montero et al. (2008) and Aleskerov et al. (2009) requires (a) random role (RR) reassignment of each subject to player  $i \in N$  in each round and (b) multiple approval (MA) of proposals that subjects can make simultaneously at a time. Under MA, all of the subject's votes are cast for each proposal he or she approves. We varied these two features; thus the experiment has

---

<sup>6</sup>A player who belongs to every winning coalition is called a veto player. It was observed that veto players eventually obtained almost everything in Montero et al. (2008), Aleskerov et al. (2009), and Esposito et al. (2012).

four treatments (protocols), as summarized in Table 1, where fixed role (FR) and single approval (SA) are also considered.<sup>7</sup> Under FR-SA, the roles of subjects are fixed during each 10 rounds and each subject can approve only one proposal, including his or her own, at a time.

In this experiment, except for there being four protocols (Table 1), in order to conduct a sensitivity analysis of the experimental results, we basically followed the experimental design of Montero et al. (2008). Each session involved 16 subjects, which were divided into two groups of 8 subjects at the beginning of each session. The subjects in each group were regrouped into two subgroups of four players at the beginning of each round. Subjects were informed ahead of time that they would be regrouped into four subgroups of four players in each round but were never informed of who else was in any particular group. Upon arrival, subjects were provided with written instructions (Appendix A), and then the experimenter read these instructions aloud. Subjects were allowed to ask questions regarding the instructions, but communication between subjects was strictly prohibited; thus, they could interact among themselves only through the information they saw via their computer screens. Subjects were informed of the fact that they would face two games, but were not informed of the apportionment of the votes nor the quota prior to starting rounds 1 and 11. Before proceeding to round 1, subjects played one practice round with a game  $[3; 1, 1, 1, 1]$  to familiarize themselves with how to use the software to play games in the experiment.

Once a round had begun, each subject could observe his or her role and other information in the upper left corner of the screen. Subject's role and the quota were highlighted in red (Figure 1). Below this information was a table that showed the number of votes each player had. In this proposal-input table, by entering into the empty cells four non-negative integers (from 0 to 100) and then pressing the "propose" button (in red), each subject could make a proposal at any time during the round. The integer entered into a cell represented the points allocated by the proposer to the player that corresponded to the cell. The sum of these integers required to be exactly the total points, which was 100. Subjects were allowed to make as many proposals as they wished during each round, but they could make only one proposal at a time. We randomized the order of the players' appearance in the proposal-input table, following Aleskerov et al. (2009), to eliminate the effect of a fixed order on the observations in this experiment.<sup>8</sup>

After a proposal was made, it was immediately shown to the other members of the relevant subgroup in proposal-approval tables that appeared on the right-hand side of the screen. Each

---

<sup>7</sup>Montero et al. (2008) and Aleskerov et al. (2009) noted the potential (positive or negative) effects of RR on how subjects learned to play the games, compared to FR, but they did not investigate experimentally how the difference between RR and FR affects subjects' behavior.

<sup>8</sup>In a game, e.g.,  $[4; 3, 2, 2]$ , if the players were ordered from left to right as players 1, 2, and 3 in the proposal-input table every round, most of the observed winning coalitions might be either  $\{1, 2\}$  or  $\{2, 3\}$ , even though  $\{1, 2\}$  and  $\{1, 3\}$  are essentially the same. Aleskerov et al. (2009) showed that randomizing the order of players' appearance in the proposal-input table every round excludes this type of effect from being observed in experiments.

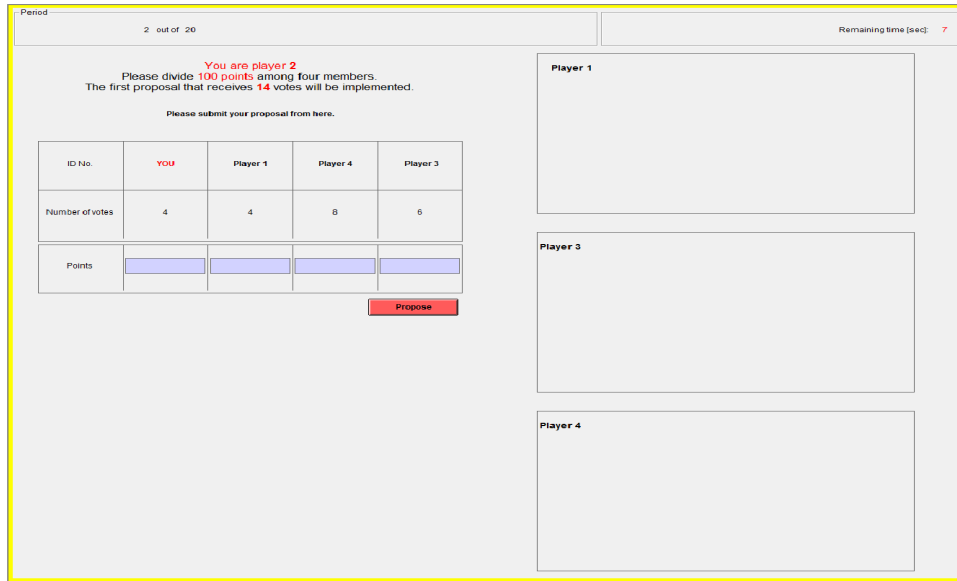


Figure 1: Screen shot for player 2. His or her proposal-input table is on the left-hand side. A proposal is made by entering four integers into the empty cells in the table and pressing the propose button (in red). The integer in a cell represents the points allocated by the proposer to the player that corresponds to the cell. The order of players' appearance in the table is randomized.

subject could approve the proposal by pressing the “approval” button (in red) in the corresponding table (Figure 2). A proposal was automatically considered to have been approved by the proposer. Making a proposal caused the proposal-input table of the proposer to be replaced by a proposal-confirmation table (Figure 3). Once a proposal was approved by a subject, a “withdraw” button (in red) appeared in his or her proposal-approval or proposal-confirmation table. At any time during the round, subjects could withdraw their current proposals or approvals by pressing this button.

Under SA, if a subject who had made a proposal wished to approve a proposal made by another, that subject had to first withdraw his or her own proposal. In the same manner, if a subject who had approved a proposal made by another wished to approve a different proposal or to make his or her own proposal, the subject had to first withdraw his or her current approval. As noted above, neither MA nor SA allowed subjects to make multiple proposals at a time. When a subject under either scheme wished to make a new proposal, he or she had to withdraw his or her previous proposal before proposing a new one.

Figure 2 shows the screen of player 2 at a time when the player has four possible actions: (1) do nothing (remain “silent”), (2) approve the proposal made by player 1, (3) approve the proposal made by player 4, or (4) make his or her own proposal to the others. When player 2 chooses (4) before any other player has taken any other action, this screen (Figure 2) will change to one of the



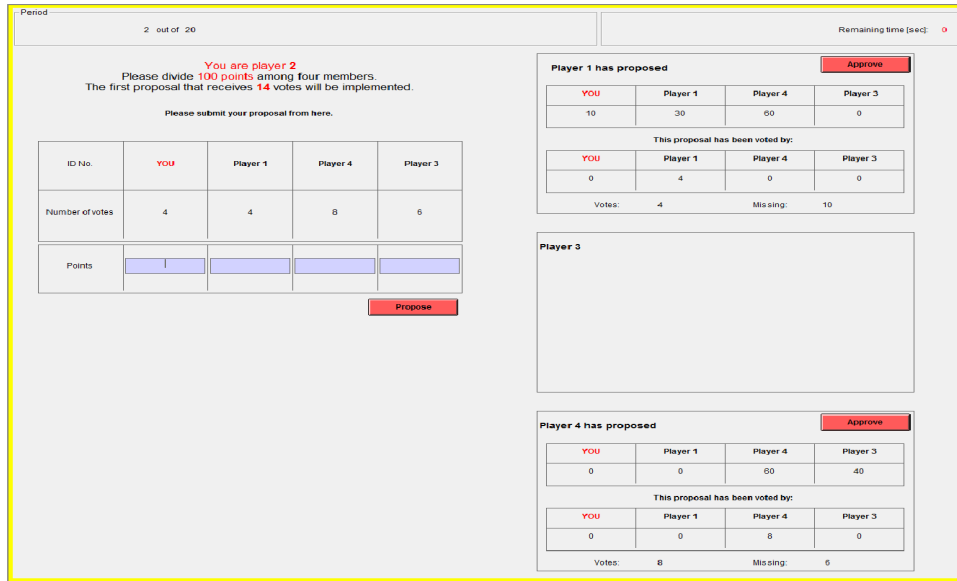


Figure 2: Screen shot for player 2, who has no current proposal; players 1 and 4 in the same group have made proposals. The proposal-approval tables of player 2 are on the right-hand side of the screen. At this point, there is no difference between the screens under MA and SA.

two shown in Figure 3. The differences in screen appearance between MA and SA can be seen in Figure 3, in which the top (bottom) figure shows the screen shot for MA (SA). Under MA, the two “approve” buttons in red on the right-hand side of the screen are still shown, in addition to the “withdrawal” button on the left-hand side of the screen, whereas they no longer appear any more under SA. Thus, SA and MA are systematically distinguished.

All the protocols shared a common time limit for subjects to reach an agreement in each round. The limit was randomly set between 300 and 420 seconds in each round.<sup>9</sup> If none of the proposals made in a group received at least as many votes as the quota within the time limit, every member of the group received zero points. Subjects were informed that they were to be paid according to the total points they gained in 6 rounds (3 from the first 10 rounds and 3 from the subsequent 10 rounds) randomly selected by a computer at the end of each experiment, following the predetermined exchange rate in addition to the show-up fee. The exchange rate was 1 point = 0.13 EUR in Montpellier and it was 1 point = 14 JPY in Tsukuba. The show-up fee was 5 EUR in Montpellier and it was 1500 JPY in Tsukuba. The instructions for each protocol is shown in Appendix A.<sup>10</sup>

<sup>9</sup>Montero et al. (2008) set the random time limit to be between 300 and 600 seconds and reported that there was no group that could not reach an agreement within 300 seconds. Aleskerov et al. (2009) set a fixed time limit of 300 seconds and reported that there were several groups that failed to reach an agreement before the time limit, but their results were otherwise almost the same as those obtained by Montero et al. (2008).

<sup>10</sup>The sentences appearing in the screens and instructions were translated into Japanese for sessions in Tsukuba and into French for those in Montpellier.

## Under MA

Period 2 out of 20
Remaining time (sec): 0

You are player 2

Please divide 100 points among four members.  
The first proposal that receives 14 votes will be implemented.

Allocation of votes.

ID No.	YOU	Player 1	Player 4	Player 3
Number of votes	4	4	8	6

My proposal Withdraw

YOU	Player 1	Player 4	Player 3
20	20	60	0

This proposal has been voted by:

YOU	Player 1	Player 4	Player 3
4	0	0	0

Votes: 4      Missing: 10

Player 1 has proposed Approve

YOU	Player 1	Player 4	Player 3
10	30	60	0

This proposal has been voted by:

YOU	Player 1	Player 4	Player 3
0	4	0	0

Votes: 4      Missing: 10

Player 3

Player 4 has proposed Approve

YOU	Player 1	Player 4	Player 3
0	0	60	40

This proposal has been voted by:

YOU	Player 1	Player 4	Player 3
0	0	8	0

Votes: 8      Missing: 6

## Under SA

Period 2 out of 20
Remaining time (sec): 0

You are player 2

Please divide 100 points among four members.  
The first proposal that receives 14 votes will be implemented.

Allocation of votes.

ID No.	YOU	Player 1	Player 4	Player 3
Number of votes	4	4	8	6

My proposal Withdraw

YOU	Player 1	Player 4	Player 3
20	20	60	0

This proposal has been voted by:

YOU	Player 1	Player 4	Player 3
4	0	0	0

Votes: 4      Missing: 10

Player 1 has proposed Approve

YOU	Player 1	Player 4	Player 3
10	30	60	0

This proposal has been voted by:

YOU	Player 1	Player 4	Player 3
0	4	0	0

Votes: 4      Missing: 10

Player 3

Player 4 has proposed Approve

YOU	Player 1	Player 4	Player 3
0	0	60	40

This proposal has been voted by:

YOU	Player 1	Player 4	Player 3
0	0	8	0

Votes: 8      Missing: 6

Figure 3: Screen shots for MA (top) and SA (bottom), after player 2 facing Figure 2 has made a proposal and no one else has taken another action yet.

### 3 Results

We conducted a computerized laboratory experiment at the University of Montpellier (France) in January 2011 and the University of Tsukuba (Japan) in October 2012. We used “z-Tree” (Fischbacher, 2007) to develop and conduct the experiment. For each of the four protocols, we had two sessions in Montpellier and two sessions in Tsukuba. At each location, game (a) was played in the first 10 rounds and game (b) was played in the second 10 rounds in one session and the order was the reverse in the other session. Each session involved 16 subjects (i.e., 256 subjects in total) and lasted about 60 minutes, including the time for administering the instructions and payments. We present the results in terms of a comparison between RR-MA and FR-SA, but also clarify which dimension, role assignment versus approval scheme, mainly accounts for the differences.

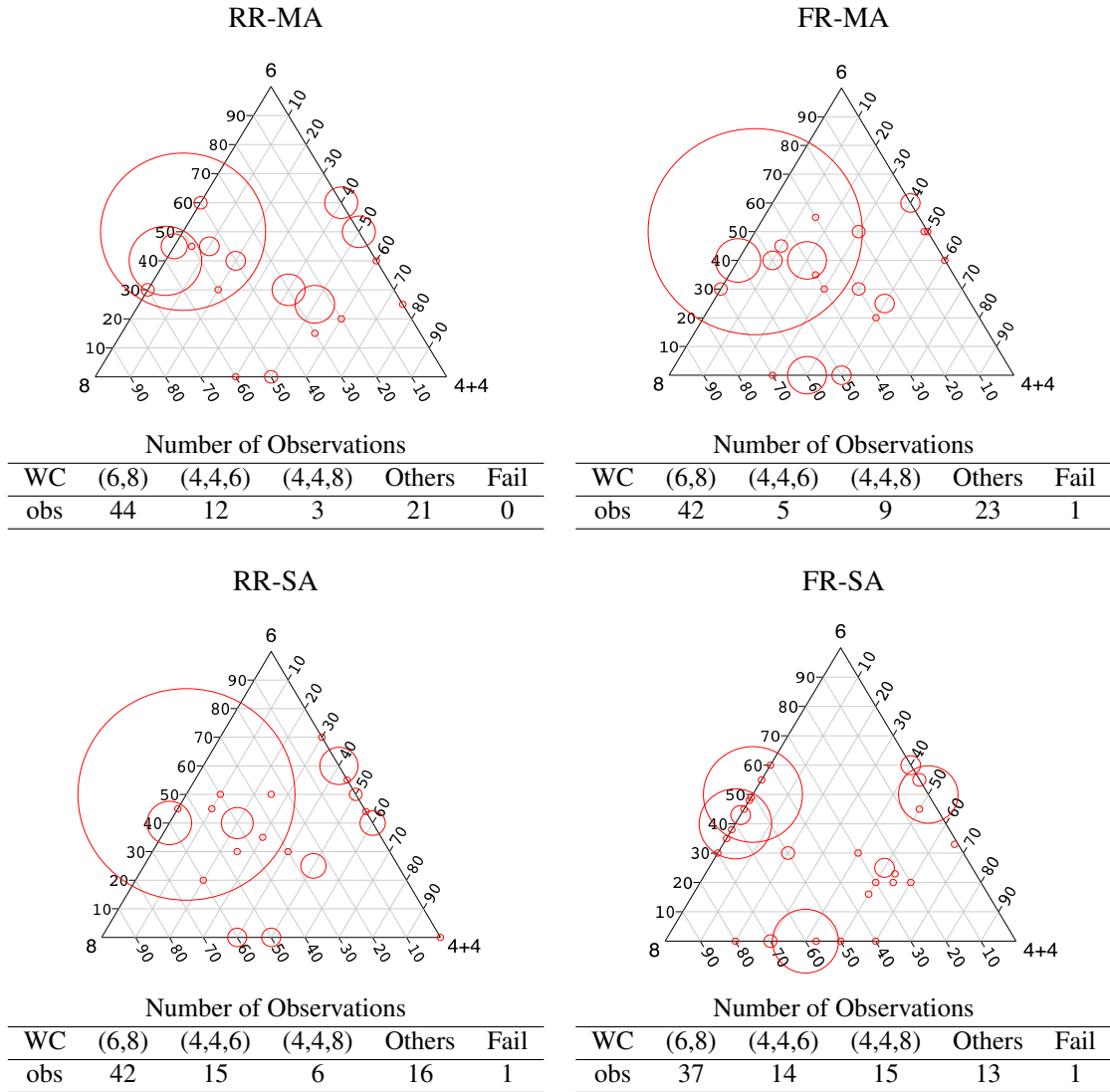
#### 3.1 Winning coalitions

In both games (a) [14; 4, 4, 6, 8] and (b) [14; 7, 7, 3, 5], there are three possible MWCs. Instead of player IDs, we hereinafter describe MWCs in terms of the number of votes the members of the coalition have. With this notation, the MWCs in game (a) are written as (4, 4, 6), (4, 4, 8), and (6, 8), while those in game (b) are (7, 7) and two MWCs of the form (7, 3, 5). Note that the two players with four votes each in game (a) and those with three or five votes in game (b) need to act together in order to be a part of a MWC.

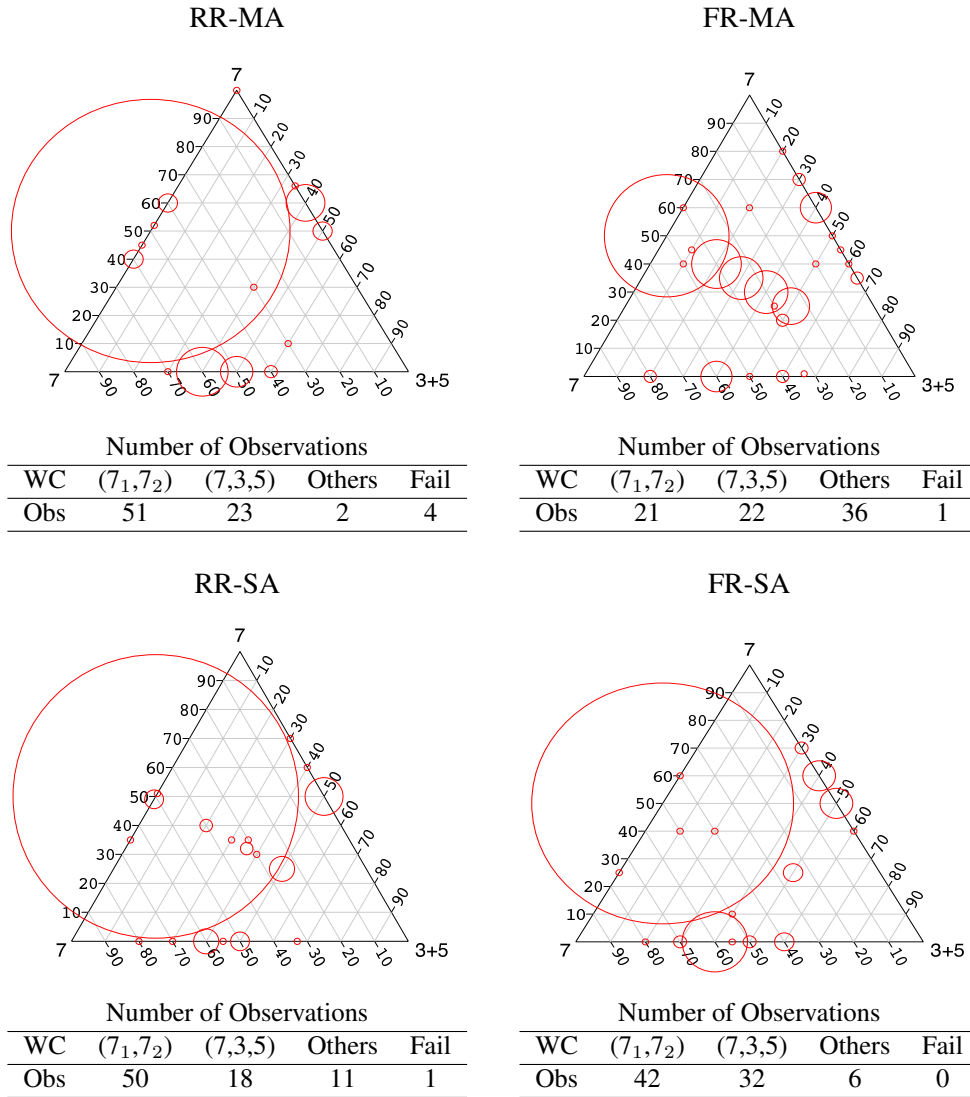
Figures 4 and 5 illustrate, in simplices, the allocations for the four subgroups observed in the last 5 rounds of games (a) and (b), respectively, played under the four protocols. The numbers of observed winning coalitions (WCs) are listed under each simplex. In each simplex, a circle that corresponds to an observation shows how the 100 points in total were allocated among the subjects and the diameter of a circle is proportional to the number of observations falling at the center of the circle. See, e.g., the simplices in Figure 4. The height of a point from the edge that is opposite from the apex labeled 4+4 represents the total points allocated for the two players with four votes each, and thus a point on the 6-8 edge represents an observed allocation that gives 0 points to both of these players with four votes, i.e., a winning coalition (6, 8) was formed. The simplices in Figure 5 should be interpreted in the same way.

**Result 1.** *In both games (a) and (b), 3-player MWCs were more frequently observed under FR-SA than under RR-MA.*

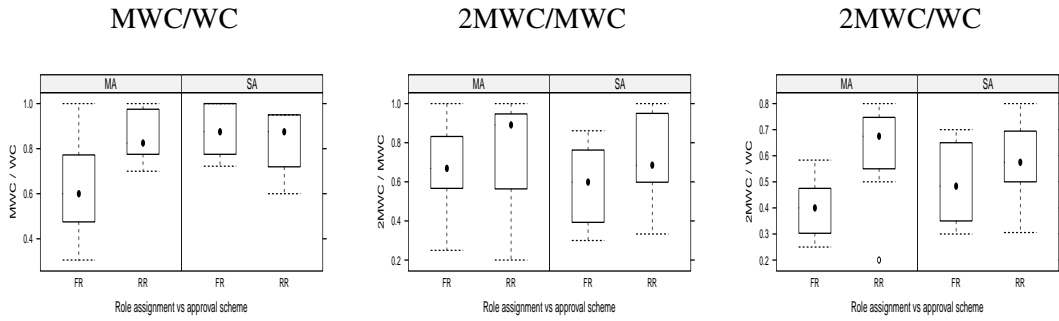
As noted in Section 1, 2-player MWCs were more likely to be observed than 3-player MWCs under RR-MA, in earlier research. We could reconfirm the difference between our earlier observations (Esposito et al., 2012) and those in the earlier research under the same game.



**Figure 4:** Distributions of allocations for four subgroups observed in the last 5 rounds of game (a)  $[14; 4, 4, 6, 8]$  played under the four protocols. The diameter of a circle is proportional to the number of observations falling at the center of the circle. The numbers of observed winning coalitions (WCs) are listed under each simplex. “Fail” means that subjects of a subgroup failed to reach an agreement before the time limit.



**Figure 5:** Distributions of allocations for four subgroups observed in the last 5 rounds of game (b) [14; 3, 5, 7, 7] played under the four protocols. The diameter of a circle is proportional to the number of observations falling at the center of the circle. The numbers of observed winning coalitions (WCs) are listed under each simplex. “Fail” means that subjects of a subgroup failed to reach an agreement before the time limit.



**Figure 6:** Boxplots of the relative frequencies of (left) MWCs as compared with all WCs (MWC/WC), (middle) 2-player MWCs as compared with all MWCs (2MWC/MWC), and (right) 2-player MWCs as compared with all WCs (2MWC/WC) observed in the last 5 rounds under the four protocols. For each protocol, we have two independent observations (two groups of 8 subjects) for in each of two sessions in each of two countries, i.e., 8 samples. The total sample size is thus 32. The minimum, lower quartile, median (filled circle), upper quartile, and maximum are depicted for each protocol.

What, then, was the most influential factor that generated different frequencies of 2-player MWCs? To find an answer to this question statistically, let us first look at a summary of descriptive statistics. The boxplots shown in Figure 6 summarize the distributions of the relative frequencies of (i) MWCs compared with all WCs (MWC/WC shown in the left panel), (ii) 2-player MWCs as compared with all MWCs (2MWC/MWC shown in the middle panel), and (iii) 2-player MWCs as compared with all WCs (2MWC/WC shown in the right panel). We consider only the last 5 rounds in each game under the four protocols to allow subjects the opportunity to learn. As noted in Section 2, the subjects were divided into two groups of 8 subjects at the beginning of each session and then further regrouped into two subgroups of four players within these 8 subjects at the beginning of each round. Thus, for various statistical tests that we report below, we treat each independent group of 8 subjects as an independent observation. We thus have two independent observations per session, and two sessions in the two countries, i.e., 8 independent observations for each treatment. Note that we pooled data that were obtained from the two games.

The left panel of Figure 6 shows that MWCs were much less frequently observed under FR-MA than under other protocols. Figures 4 and 5 show that this difference in the frequencies of MWCs is mainly coming from game (b). If we focus on MWCs, the middle panel of Figure 6 indicates that the difference in role assignments generated larger differences in relative frequencies of 2-player MWCs than the difference in approval schemes. The relative frequencies of 2-player MWCs among the observed MWCs tend to decrease in FRs. In addition, for each role assignment, the median 2MWC/MWC tends to be lower under SA than under MA. The right panel shows the frequencies of 2-player MWCs over all the winning coalitions. The relative frequencies of 2-player MWCs is

Table 2: Results of a two-way ANOVA with mixed effects.

			MWC/WC		2MWC/MWC		2MWC/WC	
	numDF	denDF	<i>F</i> -value	<i>p</i> -value	<i>F</i> -value	<i>p</i> -value	<i>F</i> -value	<i>p</i> -value
(Intercept)	1	27	156.515	<0.0001	99.078	<0.0001	508.9437	<0.0001
factor FR	1	27	1.299	0.2645	2.480	0.127	7.5035	<b>0.0108</b>
factor SA	1	27	3.207	<b>0.0845</b>	0.510	0.481	0.2968	0.5904
FRSA	1	27	7.247	<b>0.0120</b>	0.277	0.603	1.4523	0.2386

NOTE: The numerator and denominator degrees of freedom (numDF and denDF), *F*-values, and *p*-values are listed. The number of observations is 32. Primary factors are role assignment (FR=1 or FR=0 for RR) and approval scheme (SA=1 or SA=0 for MA). A possible interaction between them (FRSA=1 if FR=1 and SA=1; otherwise FRSA=0) are taken into account. The adjustment for stratification factors is made with the country (Japan=1 or Japan=0 for France). A boldfaced value indicates the rejection of the null hypothesis at the 10% significance level. The null hypothesis is that different role assignments (approval schemes) generate no difference in (a) MWC/WC, (b) 2MWC/MWC, and (c) 2MWC/WC.

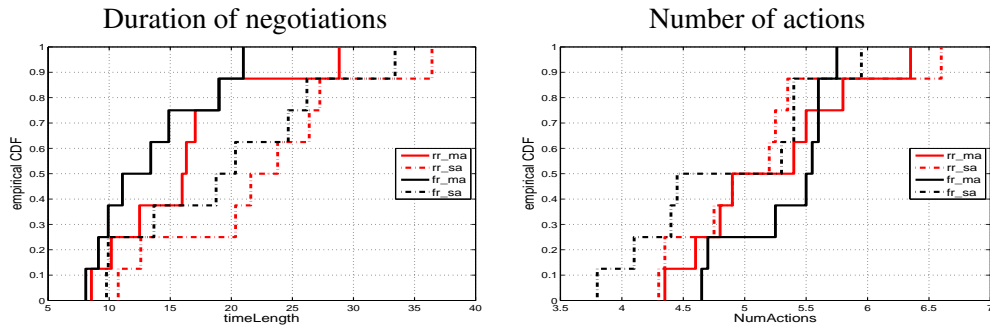
remarkably low in FR-MA but this is because the frequencies of MWCs in general are low under FR-MA.

We adopted a two-way ANOVA with mixed effects to analyze the data. We considered role assignment (FR=1 or FR=0 for RR) and approval scheme (SA=1 or SA=0 for MA) as primary factors and took into account a possible interaction between these two factors (FRSA=1 if FR=1 and SA=1; otherwise FRSA=0). The adjustment for stratification factors was made with the country (Japan=1 or Japan=0 for France). The arcsine-transformed values of the square root of the respective variables were used in the analysis according to the classical procedure (Rao, 1960; Winer, Brown and Michels, 1971). This data transformation ensured the validity of our analyses.<sup>11</sup>

Table 2 shows the results of the two-way ANOVA. The null hypothesis is that different role assignments (approval schemes) generate no difference in the values of MWC/WC, 2MWC/MWC, or 2MWC/WC. For MWC/WC, we rejected the null hypothesis on the approval scheme at the 10% significance level and the null hypothesis on the interaction of two factors at the 5% significance level. This agrees with the differences shown in the left panel of Figure 6. For the 2MWC/MWC, we could not reject the null hypothesis for either role assignment or approval scheme. For the 2MWC/WC, we rejected the null hypothesis for role assignment at the 5% significant level. These results suggest that both role assignments and approval schemes have a significant effect on the way coalitions are formed.

Why do the role assignments and approval scheme generate different outcomes in terms of relative frequencies of winning coalitions? To obtain an answer to this question, we next investigated how the protocols affected the subjects' behavior observed during negotiations.

<sup>11</sup> Appendix B shows the test results for (i) normality of sample (Shapiro-Wilk test), (ii) homogeneity of variance of samples (Levene test), and (iii) normality of residuals (Shapiro-Wilk test) conducted before the two-way ANOVA.



Result of a two-way ANOVA with mixed effects.

	$F$ -value	$p$ -value		$F$ -value	$p$ -value
(Intercept)	61.9433	<0.0001	(Intercept)	1930.247	<0.0001
factor FR	1.3428	0.2567	factor FR	0.0719	0.7906
factor SA	7.0051	<b>0.0134</b>	factor SA	1.6576	0.2089
FRSA	0.0002	0.9883	FRSA	0.5640	0.4591

Figure 7: Distribution of the average length of time (in seconds) that elapsed of the negotiation and the average number of subject actions (proposals and approvals) observed in the last 5 rounds under the four protocols. RR-MA (solid gray), FR-MA (solid black), RR-SA (dashed gray), FR-SA (dashed black). The result of two-way ANOVA with mixed effects are also shown. Primary factors are role assignment and approval scheme. A possible interaction between them are taken into account. The adjustment for stratification factors is made with the country (Japan=1 or Japan=0 for France). A boldfaced value indicates the rejection of the null hypothesis at the 10% significance level. No data transformation is carried out. The degrees of freedom for the numerators and denominators are 1 and 27, respectively.

## 3.2 How negotiations proceed

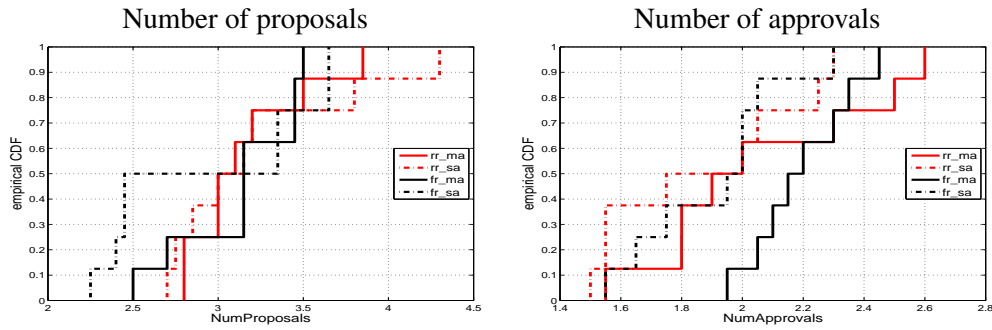
This subsection sheds light on the differences in subject behavior across the four protocols that might lead to the different outcomes shown above. We took the average within each independent observation, of which there are 8 for each protocol. Let us start with the durations of negotiation and the number of subject actions under each protocol, and then deal with the frequencies of mistakes subjects made and the proposal-objection dynamics in negotiations.

### 3.2.1 Duration of negotiation and number of actions

Figure 7 shows the distribution of the average length of time (in seconds) that elapsed of negotiations and the average number of subject actions (proposals and approvals) observed in the last 5 rounds of each game under the four protocols. In the figure, RR versus FR is represented by gray versus black lines, and SA versus MA is represented by dashed versus solid lines. The results of a two-way ANOVA with mixed effects are also shown there. No data transformation (square root followed by arcsine) is carried out for this ANOVA.

It is clear from the left panel of Figure 7 that the duration of negotiation was much shorter under





Result of a two-way ANOVA with mixed effects.

	$F$ -value	$p$ -value		$F$ -value	$p$ -value
(Intercept)	652.2150	<0.0001	(Intercept)	1563.7658	<0.0001
factor FR	0.7924	0.3812	factor FR	0.7436	0.3961
factor SA	0.1376	0.7136	factor SA	5.4786	<b>0.0269</b>
FRSA	0.5503	0.4646	FRSA	0.2428	0.6262

Figure 8: Distribution of the average number of proposals (left) and approvals (right) in the last 5 rounds under the four protocols. RR-MA (solid gray), FR-MA (solid black), RR-SA (dashed gray), FR-SA (dashed black). The results of a two-way ANOVA with mixed effects are also shown. Primary factors are role assignment and approval scheme. A possible interaction between them are taken into account. The adjustment for stratification factors is made with the country (Japan=1 or Japan=0 for France). A boldfaced value indicates the rejection of the null hypothesis at the 10% significance level. No data transformation is carried out. The degrees of freedom for the numerators and denominators are 1 and 27, respectively.

MA (solid lines) than under SA (dashed lines). The  $p$ -value is 0.0134 in the two-way ANOVA with the mixed effect for factor SA. On the other hand, the numbers of actions are not significantly different between the four protocols.<sup>12</sup> Because one can easily imagine that the number of approvals can be larger under MA than under SA, we have separated actions into proposals and approvals and plotted the distribution of the number of proposals and the number of approvals in Figure 8. As we expected, the number of approvals was higher under MA than under SA.<sup>13</sup> There was no clear difference, however, between the average number of proposals across the four protocols.<sup>14</sup>

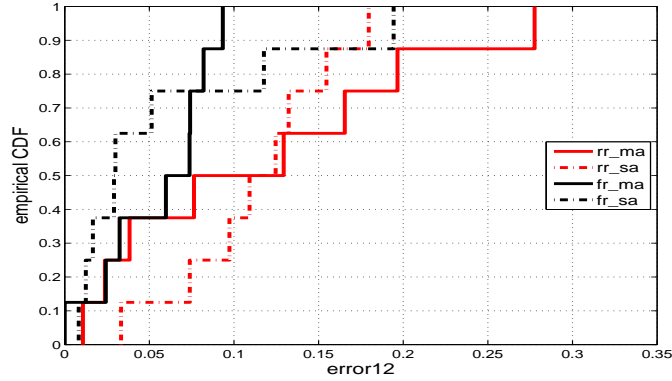
**Result 2.** *Negotiations were much shorter under RR-MA than under FR-SA.*

How does the shorter duration of negotiation, for a similar number of actions, influence subject behavior in the negotiations? An informal post-experiment comment by subjects who participated in the session for RR-MA is remarkably informative: “The negotiation processes were too fast to think and to compare the several proposals shown on the screen.” Is this just a subjective impression of the session for RR-MA by these subjects, or can we see any evidence of this in our data?

<sup>12</sup>The  $p$ -values are 0.7906 and 0.2089 for the factors FR and SA, respectively, as shown in Figure 7.

<sup>13</sup>The  $p$ -values are 0.3961 and 0.0269 for the factors FR and SA, respectively, as shown in Figure 8.

<sup>14</sup>The  $p$ -values are 0.3812 and 0.7136 for the factors FR and SA, respectively, as shown in Figure 8.



Result of a two-way ANOVA with mixed effects.

	<i>F</i> -value	<i>p</i> -value
(Intercept)	20.8896	<0.0001
factor FR	8.9657	<b>0.0058</b>
factor SA	0.0465	0.8308
FRSA	0.0857	0.7720

Figure 9: Distribution of the average normalized frequencies or errors (both type 1 and 2) in the last 5 rounds under the four protocols. The results of a two-way ANOVA with mixed effects are also shown. Primary factors are role assignment and approval scheme. A possible interaction between them are taken into account. The adjustment for stratification factors is made with the country (Japan=1 or Japan=0 for France). A boldfaced value indicates the rejection of the null hypothesis at the 10% significance level. Data transformation (square root followed by arcsine) is carried out. The degrees of freedom for the numerators and denominators are 1 and 27, respectively.

There are two implications of a short negotiation time during which intensive activities are performed, such that subjects are under a high time pressure: (i) such a time pressure can result in subjects making mistakes more frequently, and (ii) subjects simply approve proposals that were already made, rather than they propose something different from existing proposals. We will consider these two issues below.

### 3.2.2 Mistakes

Do subjects make more mistakes under a high time pressure? We observed in our data that subjects sometimes made or approved proposals that give zero points to themselves and that subjects occasionally made or approved proposals that cannot be implemented unless those who were given zero points accept such allocations. Let us call the first and the second type of mistakes errors of type 1 and type 2, respectively. Below, we normalize the frequencies of errors of the two types by the number of actions during the negotiation. The question is whether we observe more errors under RR-MA than under FR-SA.

Figure 9 shows the distribution of the normalized frequencies of errors of types 1 and 2 under

the four protocols. As one can see from the figure, the answer is mixed. The frequencies of errors are smallest at FR-MA. Errors then increase in the order of FR-SA, RR-SA, and RR-MA. It is rather easy to understand why we observe these errors much more frequently in RR than in FR (keeping approval scheme intact). The p-value for the factor FR is 0.0058 in the two-way ANOVA with mixed effects. The arcsine-transformed values of the square root of the normalized frequencies of errors are used to ensure the normalities of the variable and residuals. Because roles of subjects are randomly changed in each round in RR, while they are fixed in FR, subjects need to process more information in RR than in FR. The higher cognitive load required in RR can result in subjects committing more errors in RR than in FR. The apparent infrequency of errors under FR-MA is due to the infrequency of MWC under this particular protocol compared to others that we have discussed above. Notice that as long as subjects are proposing the grand coalition, i.e., to give positive points to everyone, errors of the both types we consider here cannot occur. The next result is due to a comparison between RR-MA and FR-SA.

**Result 3.** *Subjects made more mistakes under RR-MA than under FR-SA.*

### 3.2.3 Proposal-objection dynamics

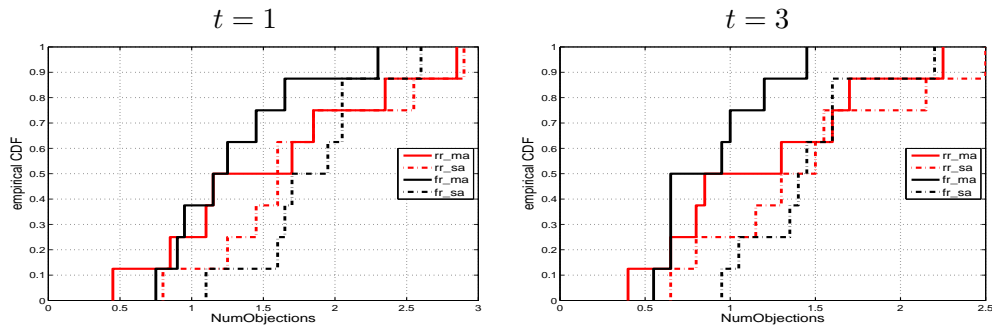
Do subjects make fewer counter-proposals in response to existing proposals under a high time pressure? We follow Aumann and Marcher (1964) in defining an “objection” to an existing proposal.<sup>15</sup> Take an existing proposal of player  $j$  such that  $j$  proposes an imputation (allocation)  $x$  with a coalition  $S$ , and write the proposal as  $(x, S)$ . We say that player  $i$  ( $\neq j$ ) has an “objection”  $(y, T)$  against  $j$  at  $(x, S)$  if (i)  $i \in T$  and  $j \notin T$ , and (ii)  $y_k > x_k$  for all  $k \in T$ . In words, an objection  $(y, T)$  made by player  $i$  against proposal  $(x, S)$  of player  $j$  must exclude  $j$  from the new coalition  $T$  (and thus give zero point to  $j$ ) and give strictly more to every member in  $T$  than what they can obtain with  $(x, S)$ .

A problem is that, when there are multiple existing proposals, it is not possible to precisely determine against which proposal a player is objecting. We thus consider all the existing proposals (i.e., those that have existed at least  $t$  seconds before the objection and have not been withdrawn at the time of objection) to be candidate proposals against which an objection is made. Did differences in protocol result in differences in terms of observed frequencies of objections?

Figure 10 shows the distribution of the average number of objections in four protocols for  $t = 1$  and  $t = 3$ . What is clear from the figure is that there are more objections observed under SA than under MA. The p-values for factor SA are 0.0596 for a delay of 1 second and 0.0212 for a delay

---

<sup>15</sup>Aumann and Marcher (1964) defined a solution concept called “bargaining set,” based on “objection” and “counter-objection.” We only employ objection and not counter-objection in our paper.



Result of a two-way ANOVA with mixed effects.

	$F$ -value	$p$ -value		$F$ -value	$p$ -value
(Intercept)	27.5089	<0.0001	(Intercept)	40.6536	<0.0001
factor FR	0.1547	0.6972	factor FR	0.8371	0.3683
factor SA	3.8663	<b>0.0596</b>	factor SA	5.9831	<b>0.0212</b>
FRSA	0.7260	0.4017	FRSA	0.8371	0.3683

Figure 10: Distribution of the average number of objections in the last 5 rounds of four treatments. Existing proposal defined as those existed at least  $t = 1$  and  $t = 3$  seconds before the time of the objection. The result of a two-way ANOVA with mixed effects are also shown. Primary factors are role assignment and approval scheme. A possible interaction between them are taken into account. The adjustment for stratification factors is made with the country (Japan=1 or Japan=0 for France). A boldfaced value indicates the rejection of the null hypothesis at the 10% significance level. No data transformation is carried out. The degrees of freedom for the numerators and denominators are 1 and 27, respectively.

of 3 seconds in the two-way ANOVA with mixed effect. No data transformation is carried out in conducting the ANOVA test.

**Result 4.** *Proposal-objection dynamics were more frequently observed under FR-SA than under RR-MA.*

## 4 Discussions

In experiments with four voters under RR-MA, the earlier research reported that MWCs of two voters were disproportionately more likely to be observed than those of three voters. The present paper showed a sensitivity of the results of weighted voting experiments with respect to changes in two features of the protocol used in the precedent research. In the two games that we examined above, MWCs of three voters were observed more frequently under FR-SA than under RR-MA (Obs. 1). Our analysis of subject behavior revealed that negotiations were much shorter under RR-MA than under FR-SA (Obs. 2). At that time, subjects commit errors more frequently under RR-MA than under FR-SA by making or accepting proposals that (i) give zero points to themselves

or (ii) can not be implemented unless those who were given zero points accept such allocations (Obs. 3), and proposal-objection dynamics were less frequently observed under RR-MA than under FR-SA (Obs. 4).

The absence of explicit structures (extensive forms) as to how negotiations take place in cooperative game environments favors MA protocols over SA protocols, although SA protocols still retain remarkably little structure as compared to the legislative bargaining model of Baron and Ferejohn (1989) and its variants (non-cooperative game environments). As Montero et al. (2008) noted, moreover, MA protocols expedite negotiations. Overly fast negotiations, however, lead to a potential problem of MA protocols that the results may be driven by subjects not making careful choices. Our experimental results thus favor FR-SA for the accumulation of data to be used in the further theoretical development of power indices. FR-SA is also preferred to RR-MA for future research on weighted voting in a cooperative game environment, because proposal-objection dynamics are more frequently observed under FR-SA than under RR-MA.

Finally, we will discuss some other observations that we intend to follow up on in our future research. Under FR-SA, in addition to the games described above, the games  $[5; 1, 2, 2, 3]$  and  $[14; 2, 6, 6, 8]$  were examined in Tsukuba. The former is the game examined in Montero et al. (2008) and Aleskerov et al. (2009). Notice that both games share the same structural property that both players who have neither the largest nor the smallest number of votes have the same number of votes. Further, the latter game has the same the same quota (14) and the total number of votes (22) as the two games we examined above. In both games and two games we examined above, there are three possible MWCs, one of which is a MWC of three voters and two of which are MWCs of two voters. The total points were 120, as in Montero et al. (2008). 16 subjects participated in each session and played the same game for 20 rounds. The exchange rate was 1 point = 10 JPY.

Figure 8 illustrates, in simplices, the allocations for four subgroups of subjects observed in the last 10 rounds. The difficulty of forming a 3-player MWC even under FR-SA was confirmed for the case that the game has the same structure as the one examined in Montero et al. (2008) and Aleskerov et al. (2009). The reason for this difficulty could be observed in the competition between 2-player MWCs, i.e., the two players with the same number of votes fiercely competed with each other in trying to form a 2-player MWC with another player by offering him or her 50-70 points. In  $[14; 4, 4, 6, 8]$  (game (a)), on the other hand, the two players with the same number of votes frequently cooperated with each other to form 3-player MWCs. Therefore, the question is, what factor in games induces subjects to compete or cooperate with each other? It would be a fruitful next step toward our final goal as stated in Section 1 to find such a factor.

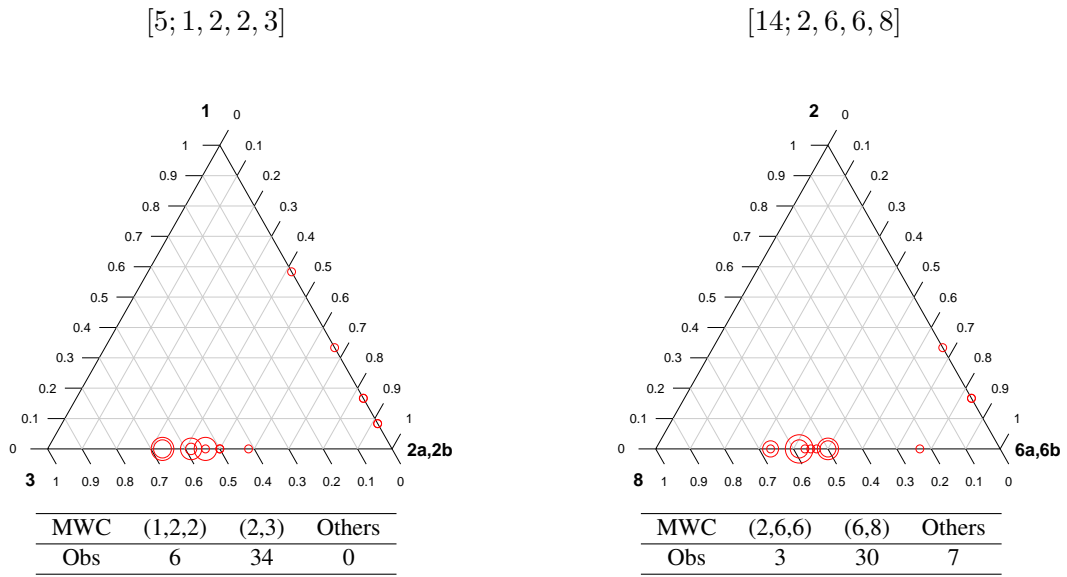


Figure 11: Distribution of allocations for four subgroups of subjects observed in game  $[5; 1, 2, 2, 3]$  and  $[14; 2, 6, 6, 8]$  from the last 10 (out of 20) rounds. The diameter of a circle is proportional to the number of observations falling at the center. The numbers of observed minimal winning coalitions (MWCs) are listed under each simplex. For the game  $[14; 2, 6, 6, 8]$ , “others” includes only  $(6, 6, 8)$ . The grand coalition was not observed in either game in the last 10 periods for these sessions.

## References

- ALESKEROV, F., A. BELIANI, AND K. POGORELSKIY (2009): “Power and preferences: an experimental approach,” mimeo., the Higher School of Economics.
- AUMANN, R. J. AND M. MASCHER (1964): “The Bargaining set for cooperative games,” *Annals of Mathematics Studies*, 52, 443–476.
- BANZHAF, J. F. (1965): “Weighted voting doesn’t work: a mathematical analysis,” *Rutgers Law Review*, 19, 317–343.
- BARON, D. P., AND J. A. FERREJOHN (1989): “Bargaining in legislatures,” *American Political Science Review*, 83, 1181–1206.
- DEEGAN, J., AND E. PACKEL (1978): “A new index of power for simple n-person games,” *International Journal of Game Theory*, 7, 113–123.
- DROUVELIS, M., M. MONTERO, AND M. SEFTON (2010): “The paradox of new members: strategic foundations and experimental evidence,” *Games and Economic Behavior*, 69, 274–292.

- ESPOSITO, G., E. GUERCI, X. LU, N. HANAKI, AND N. WATANABE (2012): “An experimental study on “meaningful learning” in weighted voting games,” mimeo., University of Tsukuba.
- FELSENTHAL, D. S., AND M. MACHOVER (1998): *The Measurement of Voting Power : Theory and Practice, Problems and Paradoxes*. Edward Elgar, London.
- FISCHBACHER, U. (2007): “z-Tree: Zurich toolbox for ready-made economic experiments,” *Experimental Economics*, 10(2), 171–178.
- FRÉCHETTE, G. R., J. H. KAGEL, AND M. MORELLI (2005a): “Gamson’s law versus non-cooperative bargaining theory,” *Games and Economic Behavior*, 51, 365–390.
- (2005b): “Nominal bargaining power, selection protocol, and discounting in legislative bargaining,” *Journal of Public Economics*, 89, 1497–1517.
- HARWELL, M. R., E. N. RUBINSTEIN, W. S. HAYES, AND C. C. OLDS. (1992): “Summarizing Monte Carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases.” *Journal of Educational Statistics* , 17: 315-339.
- KAGEL, J. H., H. SUNG, AND E. WINTER (2010): “Veto power in committees: an experimental study,” *Experimental Economics*, 13, 167–188.
- LIX, L. M., J. C. KESELMAN, AND H. J. KESELMAN (1996): “Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test.” *Review of Educational Research*, 66: 579-619.
- MONTERO, M., M. SEFTON, AND P. ZHANG (2008): “Enlargement and the balance of power: an experimental study,” *Social Choice and Welfare*, 30(1), 69–87.
- MORELLI, M. (1999): “Demand competition and policy compromise in legislative bargaining,” *American Political Science Review*, 93, 809–820.
- PACKEL, E. W., AND J. DEEGAN (1980): “An axiomated family of power indices for simple n-person games,” *Public Choice*, 35, 229–239.
- RAO, M. M. (1960): “Some asymptotic results on transformations in the analysis of variance,” *Aerospace Research Laboratory Technical Note*, 60–126.
- RIKER, W. H. (1962): *The Theory of Political Coalitions*. Yale University Press, New Haven.
- SHAPLEY, L. S., AND M. SHUBIK (1954): “A method for evaluating the distribution of power in a committee system,” *American Political Science Review*, 48, 787–792.

B. J. WINER, D. R. BROWN AND K. M. MICHELS (1971): *Statistical Principles in Experimental Design*. McGraw-Hill, New York.

## **Appendix A: Instructions**

The instructions for the four protocols are, for the most part, identical. Thus, in this appendix, we show generic instructions while noting the parts that differ between the protocols. The subjects in a particular protocol are only instructed about the relevant protocol. These instructions were translated into Japanese and French for our experiment in Tsukuba and Montpellier. Both Japanese and French versions of these instructions are available upon request.

### **INSTRUCTIONS OF THE EXPERIMENT**

Welcome! Thank you very much for taking part in our laboratory experiment.

You are a participant in an experiment on group decision making. During the experiment, you, as well as other participants in this room, will be making decisions. The experiment will take about two hours.

### **RECOMMENDATION**

We ask you to comply with these rules and respect the instructions of the experimenter. Any communication with other participants is strictly prohibited. During the experiment, you must not talk, exchange notes, watch other participants' actions, or use mobile phones. It is important that during the experiment you remain SILENT. If you have any questions, or need assistance of any kind, RAISE YOUR HAND but DO NOT SPEAK. We expect and appreciate your cooperation.

### **PROTOCOL**

There are 20 rounds in this experiment. In each round, you and three other randomly chosen participants will form a group of four people and the four players will decide how to divide 100 points in the manner described later.

### **Matching**

===== ONLY FOR THE FR =====

At the beginning of the experiment, the computer will randomly assign you a player ID number, either 1, 2, 3, or 4. Your player ID number will remain the same throughout the experiment, and will be shown on the computer screen.



At the beginning of each round, the computer will randomly group four participants with different player ID numbers into one group. You will not be able to know which participants are in the same group.

You will repeat the same procedure for 20 rounds. Your ID number does not change from round to round, but the other people in your group change.

===== ONLY FOR THE RR =====

At the beginning of each round, the computer will randomly group four participants into one group. You will not be able to know which participants are in the same group. You will repeat the same procedure for 20 rounds, but your ID number, either 1, 2, 3 or 4, may change from round to round, and the other people in your group will also change. In each round, you will be clearly informed of your player ID for that round.

=====

### **The negotiation**

You will be making a decision in a group with three other people on how to divide 100 points among the four of you.

You will not know who the people in your groups are, and the people in your group will change randomly every round.

Each player has a certain number of votes and the information will be shown in the table on the left-hand side of the screen. In the first 10 rounds, the same vote allocation will be used, and a different vote allocation will be used from round 11 to the end.

Any member of the group at any moment during the negotiation may make a public proposal about how to divide the 100 points. To make a proposal, you need to enter four numbers (integers) in the respective boxes on the left-hand side of the screen and press the “propose” button shown in red.

Any member of a group can also vote for already submitted proposals. Proposals made by others are shown on the right-hand side of the screen. You can vote for a proposal by pressing an “approve” button shown in red.

===== ONLY FOR MA =====

You can use your votes to support more than one proposal. Each proposal you support will receive all your votes. For each proposal, who are supporting it will be clearly shown.

To submit a new proposal, you need to withdraw your current proposal. You can withdraw your proposal by pressing the “withdraw” button on the left-hand side of your screen at any time during the negotiation. All the members in your group will be informed about the withdrawal of your proposal.

You can also withdraw your votes for another’s proposal by pressing the “withdraw” button shown on the right side-hand of the screen at any time during the negotiation.

===== ONLY FOR SA =====

Please remember, you can only be in favor of at most one proposal, including your submitted proposal, at any given time. You cannot divide your votes up and support multiple proposals. All your votes will be cast for the proposal that you decide to support. For each proposal, who are supporting it will be clearly shown. You can change your approval whenever you want during the negotiation.

You can withdraw your proposal in order to propose a new one or to vote for another’s proposal by pressing the “withdraw” button on the left-hand side of your screen. All the members in your group will be informed about the withdrawal of your proposal.

You can also withdraw your vote for another’s proposal to propose or to vote for a different proposal by pressing the “withdraw” button shown on the right-hand side of the screen.

=====

The first proposal that receives the necessary number of votes will be implemented and the negotiation will end. Each of your group members will receive the number of points specified in that proposal.

There is a time limit to the negotiation. The time limit will be between 300 and 420 seconds. In each round, before the start of the negotiation, the computer will randomly set the time limit, and you will not be informed of the exact time limit. This means that the round could end suddenly at any time between 300 seconds and 420 seconds after its start. If none of the proposals has received the necessary number of votes within this time limit, then all the members of your group will receive 0 points in this round.

If you have any questions, please raise your hand.

## **PAYMENT**

At the end of the experiment, the computer will randomly select 3 rounds out of the first 10 rounds and 3 rounds out of the second 10 rounds. You will be paid only according to the points you

have obtained in these selected rounds, and not according to the points of the whole protocol. The total points you have earned in the selected 6 rounds will be converted to cash at the exchange rate of 1 point = 14 JPY (13 cents in Euro).

In addition to this, you will be paid 1500 JPY (5 EUR) as a show-up fee. The maximum earning you can make is, therefore,  $1500 + 0.14 \times 6 \times 100 \text{ JPY} = 9900 \text{ JPY}$  ( $5 + 0.13 \times 6 \times 100 = 78$  EUR). The minimum earnings you can make is the show-up fee of 1500 JPY (5 EUR).

### **PRACTICE ROUND**

In order to make you familiar with the interface and mechanism of the experiment, we now do 1 practice round. What you will do in the practice will not affect your final payment. The number of votes given to the four members of your group is not related to what you will see in the real experiment to follow.

**IF YOU HAVE ANY QUESTIONS, PLEASE RAISE YOUR HAND.**

## **Appendix B: Background tests for ANOVA**

We adopted two-way ANOVA models with mixed effects to test which factor, role assignment (FR-RR) versus approval scheme (MA-SA), mainly accounts for the difference in observations. There are several assumptions to assure a correct ANOVA modeling. We tested (i) normality of samples, (ii) homogeneity of variance of samples, and (iii) normality of residuals. For some of the variables, the data transformation (square root followed by arcsine) were used according to the classical procedure (Rao, 1960; Winer, Brown and Michels, 1971).

Table 3 shows the results of the Shapiro-Wilk tests for normality of samples. Note that ANOVA is not very sensitive to moderate deviations from the normality of samples. Simulation studies have shown that the false positive rate is not affected too much by violation of the normality assumption (Harwell, Rubinstein, Hayes and Olds, 1992; Lix, Keselman and Keselman, 1996). Table 4 shows the results of the Levene tests and the Shapiro-Wilk test for normality of residuals. Please note that we need the null hypotheses to be *not rejected* for the validity of ANOVA.

Table 3: P-values for the Shapiro-Wilk test for normality of samples

	Shapiro-Wilk			
	RRMA	FRMA	RRSA	FRSA
$asin(\sqrt{MWC/WC})$	0.11	0.41	0.18	0.05
$asin(\sqrt{2MWC/MWC})$	0.53	0.86	0.32	0.76
$asin(\sqrt{2MWC/WC})$	0.10	0.63	0.95	0.28
Time Length	0.68	0.41	0.81	0.36
Number of Actions	0.48	0.04	0.21	0.84
Number of Approvals	0.75	0.99	0.22	0.54
Number of Proposals	0.04	0.14	0.12	0.22
Number of Objections 1	0.82	0.40	0.63	0.89
Number of Objections 3	0.49	0.23	0.74	0.72
$asin(\sqrt{Errors})$	0.14	0.07	0.77	0.83

NOTE: In the Shapiro-Wilk tests, the null hypothesis is that the samples are normally distributed. Objection  $t$  stands for the objection made at least  $t$  seconds after a proposal was made.

Table 4: P-values for the Levene test and the Shapiro-Wilk test for normality of residuals.

	Levene Test	Shapiro-Wilk Test
$asin(\sqrt{MWC/WC})$	0.85	0.35
$asin(\sqrt{2MWC/MWC})$	0.83	0.50
$asin(\sqrt{2MWC/WC})$	0.87	0.38
Time Length	0.49	0.54
Number of Actions	0.27	0.29
Number of Approvals	0.15	0.55
Number of Proposals	0.12	0.44
Number of Objections 1	0.35	0.26
Number of Objections 3	0.18	0.11
$asin(\sqrt{Errors})$	0.26	0.59

NOTE: In the Levene tests, the null hypothesis is that the variances of samples are equal among four treatments. In the Shapiro-Wilk tests, the null hypothesis is that the residuals are normally distributed. Objection  $t$  stands for the objection made at least  $t$  seconds after a proposal was made.