Trust in the Executive: Requiring Consensus and Turn-Taking in the Experimental Lab

T. Clark Durant*
Michael Weintraub†
Daniel Houser‡
Shuwen Li§

February 3, 2016

How can polities induce opposing elites to work together rather than seek partisan gains and political survival? This article presents a theoretical account of why the task is particularly difficult given a common feature of executive politics - the "time horizon trilemma" - that pits the need for credible commitments to trust against the need for agility in the short-term and adaptability in the long-term. Standard approaches, including winner-take-all and power-sharing institutions, force tradeoffs that do not allow opposing elites to overcome the trilemma. Inspired by executive institutions under the Roman Republic, we introduce a novel "turn-taking" institution, which requires opposing executives to take short alternating turns in office within one term. Behavioral games from the experimental lab demonstrate the trilemma's relevance as well as turn-taking's promise in overcoming barriers to agile responses, trust among elites, and adaptability of agreements.

---

*McKinsey & Company.

†Binghamton University (SUNY).

‡George Mason University.

§George Mason University. The authors thank Daniel Kling for helpful assistance in the lab, and Chris Sullivan, Mike Munger, and Ryan Oprea for useful comments. The opinions expressed in this article are solely those of the authors and do not necessarily reflect the views of those acknowledged here, McKinsey & Company, Binghamton University, or George Mason University. All errors that remain are the authors' alone. This study received Human Subjects approval from the George Mason University Institutional Review Board.

Introduction

> "Moreover you may reckon the beginning of liberty [in Rome] as proceeding rather from the limitation of the consuls' authority to a year than from any diminution of their power compared with that which the kings had exercised. All the rights of the kings and all their insignia were possessed by the earliest consuls; only one thing was guarded against - that the terror they inspired should not be doubled by permitting both to have the rods." Livy ii.1.7

How can polities induce opposing elites to work together rather than to seek partisan gains and political survival? This is a relevant question for many, whether those concerned with hyper-partisanship in established democracies (Pildes 2011; Haidt 2013; Sunstein forthcoming) or with ethnic conflict in developing countries (Walter 2002; Flores and Nooruddin 2012). Recent work has shown both why fostering elite trust and trustworthiness matters, and why this trust is so hard to sustain.[1]

Trust among elites matters because it leads to both better institutions and better policies. On the institutional side, trust can sustain constraints on the executive, preventing the executive from adjusting the rights and obligations of other citizens to punish or privilege them for political reasons (Acemoglu and Johnson 2005; North, Wallis and Weingast 2009). On the policy side, trust can allow opposing groups to harness their differences to consider a broader range of problems and solutions (Page 2007; Haidt 2013), as well as to better overcome social dilemmas in which individual and social interests diverge (Ostrom and Walker 2005). A variety of complex issues, such as sustaining economic growth, conserving the environment, improving the fairness and effectiveness of the social safety net, investing in infrastructure, require patient and concerted leadership year after year, decade after decade. This is only possible if elites can count on one another to sacrifice short term personal or political expediency for broader gains. We refer

[1] It has also examined the causes and consequences of trust in the government among ordinary citizens (Keele 2007; Hetherington and Husser 2012).

to the difficulty of getting opposing elites to trust and be trustworthy in governing together, rather than seeking short-term partisan gains and political survival, as the *chronic challenge*.

The chronic challenge is not the whole story, however; it is flanked by two others that make it even harder to address. Polities face *acute challenges* in the short term and *constitutive challenges* in the long term. Acute challenges include foreign policy threats, financial crises, and public health risks, all requiring agile responses (Moe and Howell 1999; Posner and Vermeule 2010). Constitutive challenges, on the other hand, consist of shifting identities, priorities, and beliefs of participants, which influence the rules of the game and who can play. To address constitutive challenges, there must be flexibility to iteratively adapt the terms of cooperation as the identities of players shift (Bednar 2008; North, Wallis and Weingast 2009). As a result, the credible commitment to be trustworthy is undermined on two sides: by the short-term demand for agility and the long-term demand for adaptability. The mix of the three challenges creates a *time horizon trilemma*.

Commonly recommended institutional alternatives tend to address one or two but not all three sides of the trilemma. Winner-take-all institutions allow agile responses by choosing one decisive leader. That there can be only one leader, however, creates high stakes competition to win or retain control, undermining trust among elites. The "shadow of the future" can do little to foster cooperation and lower the stakes if vanquished candidates have at best distant and uncertain prospects of competing to win again (Fearon 1995; Weingast 1997; Robinson 1998; Bueno de Mesquita and Smith 2010). High stakes competition creates pressure to pursue "bad policy" so long as it is "good politics" (Robinson 1998).[2] In the extreme, the *ex ante* prospect or *ex post* reality of losing

[2]This can mean targeting short-term benefits (e.g, tax breaks, subsidies, monopoly privileges) at a potentially pivotal group at the expense of long-term benefits for the polity as a whole (Rodrik,

a high stakes election can lead to coups, riots, or civil wars (Bates 2008).

Power-sharing or power-dividing institutions, on the other hand, allow for credible commitments by creating multiple offices, coalition members, or branches that have a veto, and stipulating that all veto-holders must agree to any change in policy (e.g. Tsebelis 2002; Lijphart 2004). This makes commitments credible, provided agreements can be made upfront, and also reduces the stakes of electoral competition for any particular office, each of which has less influence over the system as a whole. At the same time, such institutions reduce agility: changes trigger bargaining behaviors, with veto-holders seeking to secure a greater share of the gains from trade. In some cases, intransigence and brinksmanship can block any adjustment at all (Goodin 1996), preventing agile responses in the short term and slowing iterative adaptation of the terms of cooperation over the long term.

To address the time horizon trilemma, a more robust approach to cooperation through time is required. Fortunately, over the last thirty years there has been an outpouring of work across political science, economics, and psychology on the potential of *reciprocity* to drive cooperation and trust through time, particularly when conditions are right. What makes the conditions right? Propitious conditions include repeated interactions between a small number of mutually identifiable players with clear rules and norms about what behaviors are obligatory, permitted, and forbidden; players who have the ability to form reputations based on mutual observation, to reward or punish one another; and the ability of players to iteratively assess and improve their institutions to match shifting circumstances.[3]

Subramanian and Trebbi 2004; Acemoglu and Johnson 2005); pursuing policies that mobilize co-partisans, bribe swing voters, and/or divide the opposition; and incentivizing elites to showcase and/or hasten the benefits of their agenda while hiding and/or delaying its costs (Cowen 2005).

[3]Initial work took a game-theoretic lens, and focused on getting *strategic structure* right (Axelrod 1984). Subsequent empirical work on real-world social dilemmas has incorporated more behavioral, social, and institutional nuance (Ostrom 2005).

This work on how to create conditions conducive to reciprocity and trust allows us to both draw inspiration from and better understand the slow and steady rise of the Roman Republic. Livy's conjecture, recounted above, was that the creation of the office of the consul in ancient Rome was pivotal in moving from the informal patron-client networks that defined elite relations under the monarchy to the formal obligations and liberties that increasingly defined them under the Republic. The institution of the consular office consisted of (a) splitting the office in two, (b) electing the pair or "college" to a one-year term, and (c) forcing turn-taking from month to month within the term so long as both consuls were in Rome (e.g. Cornell 1995). While likely intended as an interim measure, this executive institution endured with minor variations for more than 400 years as Rome transformed itself from a provincial city-state to a Mediterranean empire.[4]

We test three hypotheses in the experimental lab related to incentivizing trust among opposing elites, inspired by the Roman case.[5] Elsewhere we use game-theoretic models to consider a modern adaptation of the Roman consulate where sufficient consensus (e.g. 55% of the electorate) yields a single winner, but a divided electorate yields turn-taking between the top two vote-getters, e.g. taking alternating annual turns twice over a four-year term. To establish a base case for use in the experimental lab, in this paper the assumption is that the electorate remains divided, such that neither side has a chance of winning the whole term, thereby requiring turn-taking. We hypothesize that the turn-taking institution should preserve agile responses in the short term (e.g., in a given year), foster mutual trust and cooperation over the medium term (e.g., between

[4]The primary variation was that - for 52 of the years over the span from 444 to 367 BCE - there were 3 to 9 consular tribunes elected instead of just two consuls (Cornell 1995). Within the basic 2-consul design, the most consequential variation was requiring a 1 to 1 split between patricians and plebeians in 342 BCE.

[5]The lab allows researchers to evaluate the comparative performance of institutions that do not exist in the real world, do not exist in sufficient numbers to be amenable to large-N analysis, or do not lend themselves to randomized controlled trials in the field.

a particular pair of leaders over the course of a few terms), and enable iterative adaptation over the long term (e.g., as different pairs of leaders take office over the course of decades). To test these propositions, we use a 2 x 2 experimental design across 4 institutional environments: a winner-take-all baseline, a power-sharing treatment ("requiring consensus"), a turn-taking treatment ("requiring turn-taking"), and the latter two combined ("requiring consensus and turn-taking").

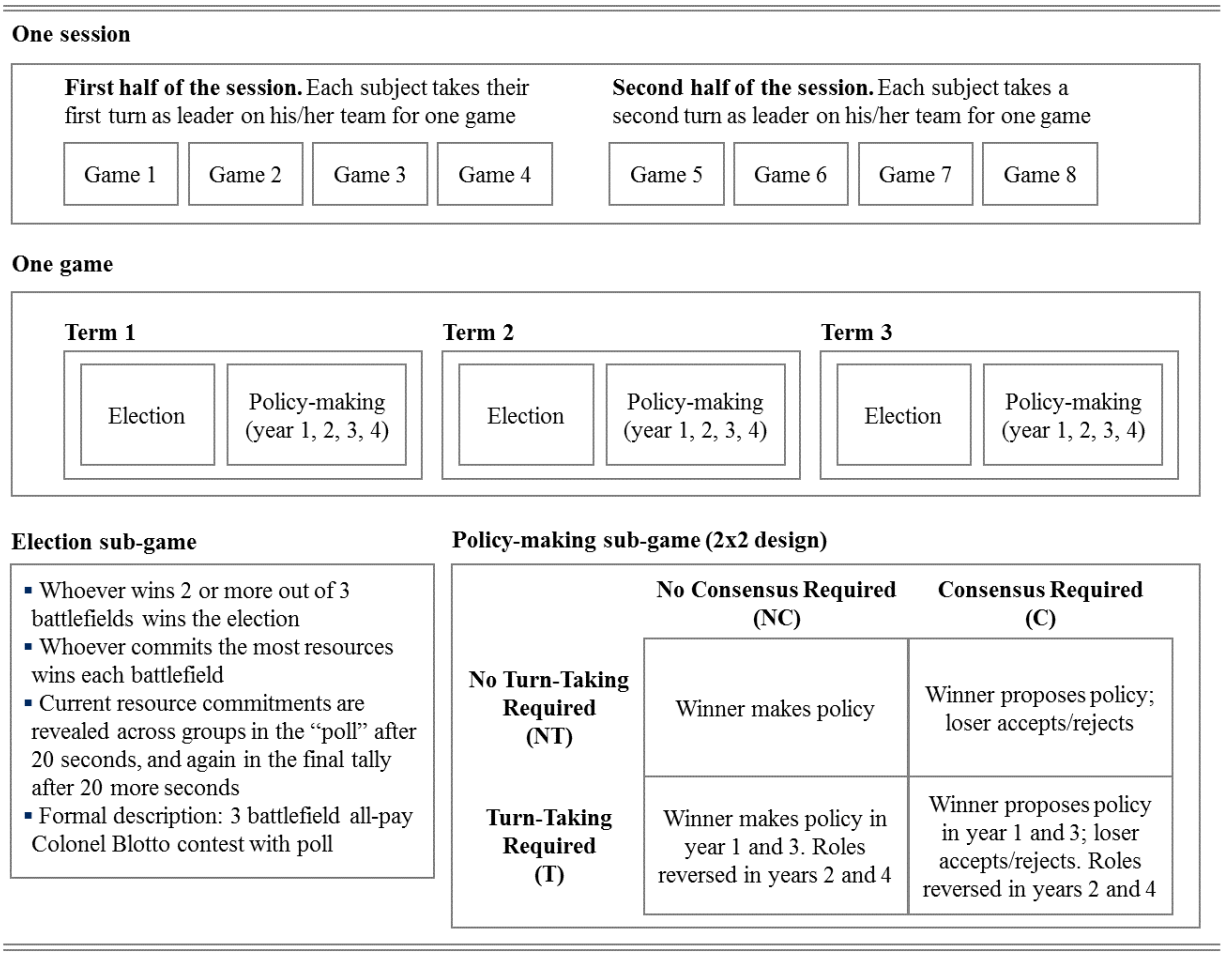
The results from the experiments validate the core features of the time horizons trilemma and demonstrate the potential of turn-taking as a countermeasure. The winner-take-all baseline allows for agility in the short term (in a given "year"), but at the cost of high stakes electoral conflict and low trust and trustworthiness among elites in the medium term (over the course of 3 four-year "terms"). "Requiring consensus" lowers the stakes and increases trust and trustworthiness, but at the cost of agility and iterative adaptability. By contrast, "requiring turn-taking" reduces agility by less than "requiring consensus" in the short term, improves trust by more in the medium term, and increases adaptability by more in the long term. In sum, turn-taking demonstrates great promise in overcoming core barriers to agile responses, trust among elites, and adaptability of agreements.

Experimental design

To assess the effects of executive institutions on agility, trust, and adaptability, we combine a standard electoral game with a policy-making game that varies according to a 2x2 experimental design. Victory in the election game provides the chance to be "in office" in the policy-making game. This reflects key features of the political environment, in which office-holders pursue public policies to help them retain office in the future, and in which spending on elections depends upon how costly elites believe it is to be out of office.

We begin with two teams, consisting of four players each. Those teams are kept constant throughout the session. In each session, subjects play a series of 24 terms, with a term consisting of an election game and a policy-making game. In every game, each team had one player assigned as the leader, who was responsible for making policy decisions during the three terms per game. Each subject was randomly assigned the leader role for a three-term "match" in the first half of the session (e.g. 3 consecutive terms in the first 12 terms), and then again for another "match" with the same counterpart in the second half. The general setup is depicted in Figure 1.

Figure 1: Experimental design



In the election game, subjects deploy their resources to win the right to make policy in the policy-making game (e.g. to be "in" office). As the focus is on elite behavior - rather than mass citizen behavior - we follow others in using a strategic "contest" rather than a voting game (Merolla, Munger and Tofias 2005; Garfinkel and Skaperdas 2007). To win the office, subjects commit resources to three "battlefields." Whichever team commits the most resources to a battlefield wins that battlefield, and whichever team wins the most battlefields wins the election.[6] Subjects were given 40 seconds to make or adjust their bids, with their choices only visible to their teammates, with one exception: resource placements at the 20 second mark were made mutually visible, so players had a chance to react and strategically manipulate their opponents' reactions. Ties were settled by a coin toss. The election game is held constant across treatments.

The policy-making game is varied, with four treatments: the baseline no consensus, no turn-taking (NC-NT); requiring consensus, no turn-taking (C-NT); no requiring consensus, requiring turn-taking (NC-T); and requiring consensus, requiring turn-taking (C-T). For both of the "requiring turn-taking" treatments, whichever team wins the election chooses a policy for the first and third years of the term, but roles are reversed for the second and fourth years. For the "not requiring turn-taking" treatments, whichever team wins the election chooses a policy for all four years of a term. For the "requiring consensus" treatments, the leader of whichever team wins the election ("in" team) proposes policy payoffs for the year to the leader of the losers ("out" team), who can accept or reject them. As in standard ultimatum game environments, rejecting means that neither side receives a payoff.

[6]This is equivalent to a 3 battlefield, all-pay Blotto contest to win an endogenous prize. On the history and dynamics of Blotto games, see Golman and Page (2009). Battlefields could be understood as geographical areas (e.g. states deemed "in play" in the Electoral College), or demographic or issue-based segments of the electorate that are undecided (e.g. soccer moms, millennials, or cultural conservatives).

Subjects began each term with an initial endowment of \$10. Payoffs at the end of the term equaled the initial \$10 endowment minus any resources committed during the election game plus any resources won from policy payoffs. Payoffs at the end of each game were reduced to average payoffs for each term. To determine the subjects' take-home winnings, one game was picked at random, with players paid depending on their performance in that game.

A total of 176 subjects participated in 528 elections and made a total of 2,112 policy choices over the course of 22 independent sessions.[7]

We now formalize the description of the election and policy-making games for greater precision and to provide additional details about the procedure.

The electoral environment: constant across treatments. Each member i, where i=1, 2, 3, or 4, of group k, where k=1 or 2, may bid any whole dollar of her endowment, e, on each of three battlefields, b=1, 2, or 3. Each individual's final bid summed across battlefields is denoted by x_{ik}, where $0 \leq x_{ik} \leq e$. The total bid by group k on battlefield b is denoted by X_{bk}, where $X_{bk} = \sum x_{ibk}$. The probability that group k wins a given battlefield b is:

[7]All subjects were volunteers recruited from the undergraduate student population at George Mason University. The computerized experimental sessions were programmed using z-Tree (Fischbacher 2007). No communication between subjects was permitted and all choices and information were transmitted via computer terminals. All decisions were anonymous. Subjects were given the instructions, available in Appendix 1, and the experimenter read the instructions aloud as subjects followed along on paper. Before the actual experiment, subjects played two practice terms and answered questions on a quiz to ensure comprehension. The experiment started only after all subjects had answered the quiz questions and explanations were provided for any incorrect answers. No subjects played more than one treatment. After running 4 sessions of each of the 4 treatments, 3 more of each of the "requiring turn-taking" treatments (T-NC and T-C) were run, as they exhibited greater variability in policy efficiency due to the greater frequency of conditional play: bursts of cooperation, then defection and punishment, then a return to cooperation, and so forth.

$$p_{bk} = \begin{Bmatrix} 1 & \text{if } X_{bk} > X_{b\sim k} \\ 0.5 & \text{if } X_{bk} = X_{b\sim k} \\ 0 & \text{if } X_{bk} < X_{b\sim k} \end{Bmatrix} \tag{1}$$

The payoff for individual i of group k in a given term equals her endowment minus her contributions to each battlefield plus the policy payoff for the term, which is v_{1st} for the 1st place group and v_{2nd} for the 2nd place group:

$$\pi_{ik} = \begin{Bmatrix} e - x_{ik} + v_{1st} & \text{if } k \text{ wins the majority of battlefields} \\ e - x_{ik} + v_{2nd} & \text{otherwise} \end{Bmatrix} \tag{2}$$

The stakes of the election, $v_{1st} - v_{2nd}$, vary according to the choices made within the four policy-making games.[8]

The policy-making environment: a 2 x 2 design. For each "year" a leader is in office, she chooses a "policy," dividing \$5 between two public goods accounts: one account for herself (and her partisans), and one for the other office-seeker (and her partisans). Because of the policy multiplier, each \$1 a leader puts into the in-party account generates \$1 of benefits for each subject in her own party, while each \$1 she puts in the out-party account is multiplied so that it generates more than \$1 for each subject outside of her party. With a 2x multiplier, the leader can keep \$5 for herself and each of her partisans, or give \$10 to those in the other party, or any dollar unit in between.[9] All policy options are shown in Table 1. This payoff structure - with each incremental dollar creating \$2

[8]See Appendix 2 for a detailed account of how we derived numerical solutions for the election game given the stakes of the policy-making games. The challenge is that there is no known closed form solution for the Colonel Blotto game with three battlefields between two groups each of four players, using dollar unit bids.

[9]The menu is a mix of a dictator's dilemma payoff structure (Cherry, Frykblom and Shogren 2002) and a public goods payoff structure.

for the out-group than $1 the in-group - is patterned after standard trust games.[10] In a political context, this reflects that (a) policy preferences differ and (b) policy-makers may not take into account all costs imposed on those outside their group.[11]

Table 1: Policy menu

Policy	Payoff for own side ("Ins")	Payoff for other side ("Outs")
A	$5	$0
B	$4	$2
C	$3	$4
D	$2	$6
E	$1	$8
F	$0	$10

The acute challenge and the level of agility. Our measure of agility, seconds per policy choice, is meant to reflect the value of agile responses in the face of shifting circumstances, both to seize hold of fleeting upside opportunities and to mitigate the downside of adverse challenges. This metric builds on a commonly used measure of agility in operational environments: lead time, the time elapsed between the request for an output and its fulfillment (Cachon and Terwiesch 2009). As we discuss below, this measure serves only as a starting point. The challenge in the real world is to address as many high priority issues as possible in a fixed amount of time in office, not a single high priority issue as fast as possible. We hypothesize that agility is a function of (a) the number of decision-makers, and (b) the strategic complexity of the choice, with the former factor carrying more weight than the latter. Thus we expect that the baseline (e.g., one decision-maker per year, without having to strategically contemplate immediate responses) will

[10] While this game appears similar to a centipede game (e.g. Rosenthal 1981), in that it is finite and, by giving more resources, more are "created," it more closely resembles standard trust games.

[11] While it would be possible to create a variant where the in-group can make these concessions on the margin *while still receiving higher payoffs in absolute terms*, we opted to keep things simple and comparable with other experiments, and to note that the marginal payoffs may be more externally valid than they at first appear.

be fastest, then requiring turn-taking only (one decision-maker per year, but having to strategically contemplate some immediate responses), then requiring consensus only, then requiring both consensus and turn-taking.

The chronic challenge and the level of trust. Cooperation over and above the subgame perfect Nash equilibrium (SPNE) in dictator and trust games is a particularly good measure of trust and/or trustworthiness (e.g. Berg and McCabe 1995; Glaeser et al. 2000; McCabe, Rigdon and Smith 2003). For the treatments where no consensus is required (either NC-NT or NC-T), the unique SPNE is to pick policy A each year.[12] For the treatments where consensus is required (either C-T or C-NT), the relevant SPNE is for one side to propose policy B, and the other side to accept it.[13] These equilibria hold in the commonly cited case of finitely repeated games with subjects (a) who are rational, (b) who are either selfish or groupish, and (c) for whom the common knowledge assumption holds.

The constitutive challenge and the level of adaptability. We measure iterative adaptability by counting the number of potential trust opportunities realized in a given session. A potential trust opportunity occurs when the "trustor" takes office. In other words, a trust opportunity is a trust game that emerges endogenously from turnover in office, based on changes in who wins elections or how elections allocate power. A trust opportunity is "realized" - and therefore an agreement is iteratively adapted - if three additional conditions are satisfied: (1) a trustee must take office after the trustor, (2) trust must be placed in that trustee, and (3) that trustee must prove trustworthy by paying back

[12]Notably, this equilibrium would be the same if subjects were selfish, seeking to maximize their individual payoffs, or "groupish", seeking to maximize payoffs to their own group.

[13]There is one other SPNE for this game: the proposer offers policy A, and the responder - indifferent between accepting and rejecting - accepts it. In practice, this makes it costless for the responder to reject the offer. The threat of this is part of what makes offers higher in the ultimatum game than in the dictator game.

the initial trust invested.[14] We hypothesize that requiring turn-taking should generate more realized trust opportunities than requiring consensus because role reversal triggers positive and negative reciprocity *even when the SPNE would be non-cooperation.*

We can summarize our hypotheses as follows:

> $H1$: Requiring turn-taking reduces the agility of policy-making by less than requiring consensus.
>
> $H2$: Requiring turn-taking increases trust and trustworthiness by more than requiring consensus.
>
> $H3$: Requiring turn-taking increases the number of realized trust opportunities by more than requiring consensus.

Experimental results

For each of our hypotheses we describe the basic results. Overall, the results are consistent with the hypotheses. The turn-taking treatments do more in the short term to preserve agility; more in the medium term to lower the stakes of electoral conflict and increase trust and trustworthiness in policy-making; and more in the long term to allow iterative adaptability and promote a greater number of trust opportunities. Figure 2 reports two-tailed t-tests for institutional comparisons. A subsequent section discusses mechanisms underlying these results and explores refinements of our hypotheses.

Our first hypothesis is that requiring turn-taking decreases the agility of policy-making by less than requiring consensus. Figure 2 shows that requiring consensus does decrease agility, with a difference-in-means significant at $p < 0.000$. Requiring turn-taking increases seconds per policy decision as well, but by too small an amount to be statistically significant. This is consistent with the hypothesis.

[14] This is a relatively low bar for determining trustworthiness, given that the policy multiplier automatically doubles what is given, reflecting the notion of the "break-even point" used in business to assess the feasibility of an investment.

Figure 2: Experimental results support core hypotheses

Hypotheses	Dependent variable	Institution A Mean (SEM)	Institution B Mean (SEM)	Δ of means (A-B)	*p*-value
Short term 1. Requiring turn-taking decreases agility by less than requiring consensus	**Agility.** Seconds per policy decision	Consensus 11.36 (1.44)	noConsensus 7.61 (1.61)	3.75	0.000
		TT 8.03 (3.36)	noTT 7.07 (5.11)	0.96	0.642
Medium term 2. Requiring turn-taking increases trust by more than requiring consensus	**Trust.** Annual policy payoffs per person over the selfish subgame perfect Nash equilibrium	Consensus $0.23 ($0.72)	noConsensus $0.41 ($0.51)	-$0.18	0.616
		TT $0.63 ($0.64)	noTT $0.01 ($0.29)	$0.65	0.005
Long term 3. Requiring turn-taking increases the number of realized trust opportunities by more than requiring consensus	**Adaptability.** Number of realized trust opportunities, i.e., trustors who trust and have a trustworthy counterpart	Consensus 12.45 (13.55)	noConsensus 8.18 (9.56)	4.27	0.404
		TT 15.14 (12.15)	noTT 1.88 (2.23)	13.27	0.001

Our second hypothesis is that requiring turn-taking increases trust by more than requiring consensus. We measure trust between groups as the annual payoffs per person above the "selfish" SPNE. Table 2 shows that requiring consensus fails to increase trust, but turn-taking does significantly increase trust, at $p < 0.001$. Hypothesis 2 finds support in the data.

Our third hypothesis is that a greater number of trust opportunities are realized per session in turn-taking institutions relative to other alternatives. Figure 2 shows that requiring consensus increases realized trust opportunities, though not at a statistically significant level, whereas requiring turn-taking increases realized trust opportunities by nearly seven times, a difference that is significant at $p < 0.001$. Hypothesis 3 finds support. Appendix 3 visually demonstrates how the number of realized trust opportunities varies across treatments.

Discussion

Here we focus on four insights we find counter-intuitive and theoretically compelling. Each of the first three pertains to one of the three hypotheses, while the fourth addresses overall motivations of subjects.

A. On agility: setting a baseline

The naïve prediction is that requiring consensus and requiring turn-taking would each reduce the agility of policy decisions, and the two together even more so. Our first hypothesis - that requiring turn-taking would reduce agility by less than requiring consensus - was a prediction on the relative size of these effects, and the results were consistent with this. "Requiring consensus" (rather than not) added 50% more time to each policy decision (e.g. four seconds) at $p < 0.001$. "Requiring turn-taking" (rather than not) added 13% more time to each policy decision (e.g. one second), though this was not statistically significant.

There are obvious pitfalls to extrapolating from our simple setup in the lab to the executive office. External validity of this result could be limited by the complexity of real policy choices, the difference between university students and experienced officeholders, and the contrast between the isolated lab set-up and the elaborate social and organizational apparatus that surrounds real officeholders.

Even with these caveats, three points are worth making. First, the agility hypothesis is the least controversial or counter-intuitive of the three. The best reason to believe that requiring turn-taking would reduce agility by less than requiring consensus does not stem from our experimental results, but the general fact that requiring consensus from more participants forces discussion and negotiation that takes more time *across a wide variety of types of situations and decision-makers*, from the trivial (e.g. "where should we

eat dinner") to the existential (e.g. "should we go to war?"). Second, these are still the right measurements to take at this early stage of inquiry. The speed of decisions made, problems solved, or results delivered is a near universal metric of agility in operational environments (Cachon and Terwiesch 2009). Turnaround time is used as a measure of agility from the most simple tasks (those that take seconds or minutes) to the most complex tasks (those that take weeks or months), and more moderate complexity tasks in between (those that take hours or days). While the task in this current set-up is a simple one, it provides a baseline point of comparison for future work could make use of more complex task environments. Finally, even with this current set-up, there is one counter-intuitive finding, namely that the increase in agility over the course of the experimental session is greater for requiring turn-taking than for requiring consensus.

Figure 3: Experimental results, focusing on agility

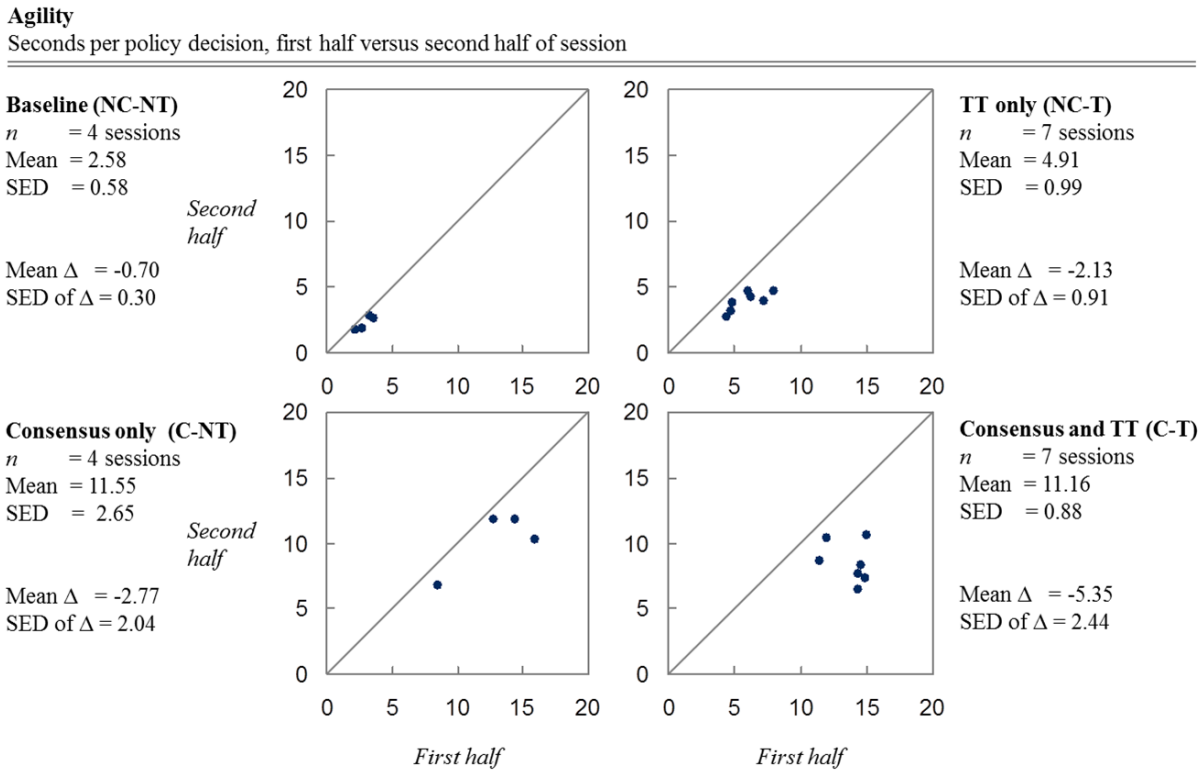


Figure 3 plots across each of the treatments the number of seconds until a policy

allocation was made, with the x-axis plotting the first half and the y-axis the second half of the experimental session. Improvements in agility would manifest themselves as data points clustered on the right side of the figure. What we see are significant improvements in the consensus only (C-NT) and the consensus with turn-taking treatments (C-TT), and more modest improvements in the turn-taking only (NC-TT) treatment. While mean seconds to decision in the baseline (NC-NT) treatment is much lower than in others, the average change in the two consensus treatments - albeit with a larger standard error of difference - indicates significant improvements. For the turn-taking treatments, the average effect is a 32% reduction, whereas for the no "turn-taking" treatments, the average effect is a 20% reduction. This 12 percentage point difference is significant at $p < 0.002$. By contrast, there is no statistically significant difference between "requiring consensus" treatments (which show a mean reduction of 29%) and "no consensus" treatments (which show a mean reduction of 27%).

Why would this be? One possibility is that the strategic complexity that subjects face is not constant, even within a given treatment. While strategic complexity likely falls as subjects gain experience, requiring turn-taking causes reaction times to fall *faster* than requiring consensus, suggesting that subjects within turn-taking treatments more rapidly approached a state of automatic or intuitive play (Logan 1992).

B. On trust: requiring consensus can trigger bargaining

The naïve prediction is that requiring consensus should foster cooperation. Counter-intuitively, however, "requiring consensus only" does the *least* to foster trust. "Requiring consensus only" only seems to make short-term bargaining behaviors more salient. In its more benign forms, this led to the minimum amount of cooperation that is incentive-compatible in the short term (e.g. a given year). Even "equitable" proposals - splits, or as close to even as possible - can be a form of destructive short-term bargaining behavior,

as this takes the most socially productive policies off the table (e.g., policies D, E, and F in Table 1). Most of the proposals (e.g. 83%) are clustered around the most equitable options, policies B and C.

The less benign form of short-term bargaining behavior is brinksmanship, which involves committing to extreme positions and risking that no agreement is reached to increase one's share of the surplus (Goodin 1996). While 37% of policy choices made in the "requiring consensus only" institution exhibit trust, e.g. policies more generous than the selfish SPNE, these are counterbalanced by the fact that 19% of proposed policies are rejected. Such rejections have a disproportionate impact on policy payoffs: moving from policy A to B increases policy payoffs by $0.50 per person on average, while rejecting policy B is equivalent to reducing policy payoffs by $3.00 per person on average. Of the 19% of proposals that were rejected, one third had proposed policy A, offering nothing to their counterpart, while two-thirds had proposed policy B, offering $2 and keeping $4 per person. Offering nothing is puzzling because it makes it costless to reject.[15]

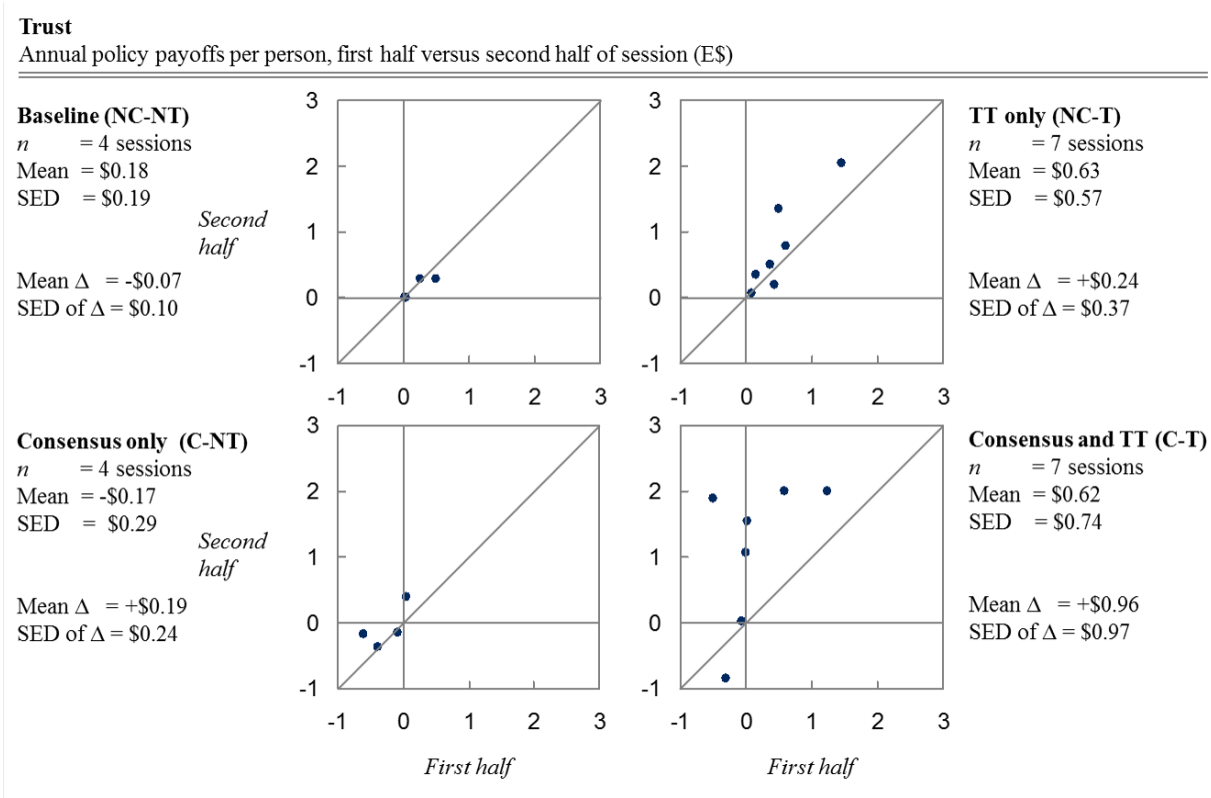
Why, then, do we see so many breakdowns? One hypothesis is that some subjects are simply confused as to the incentives. Yet all subjects played practice rounds, took a quiz to test their understanding, and all had the opportunity to learn over 24 terms (with 6 terms played as the leader). Another, more plausible hypothesis is that proposers make lower offers, and responders accept lower offers, when proposers have "won the right" to their privileged role (Hoffman et al. 1994). This could apply in our case insofar as proposers have just won an election. If this were true but not the former, we would expect to see (a) some responders accept policy A in the "requiring consensus only" institution, and (b) the officeholders in the winner-take-all baseline institution offer less than those in otherwise comparable dictator game environments. In fact, responders accept 40% of the

[15]In standard ultimatum games, it is commonplace for proposers to offer 40% of a fixed endowment on average, and for responders to accept anything above 20% (e.g. Slonim and Roth 1998).

occasions when policy A is proposed. Moreover, whereas 60% of dictators in standard environments give some positive amount, with the mean giving approximately 20% of their endowment to their counterparts (e.g. List 2007) (approximating choosing policy B in our environment), in our baseline institution we see giving in only 19% of the time. Both points suggest proposers will offer and responders will accept lower payoffs when proposers feel they have "earned the right" to their privileged position.

Here too we see a divergence in subject play as the game progresses, shown in Figure 4. Trust improves most in the consensus, turn-taking (C-T) and turn-taking only (NC-T) treatments. The baseline (NC-NT) and the requiring consensus only (C-NT) treatments, on the other hand, show no statistically significant improvement in trust across time.

Figure 4: Experimental results, focusing on trust



C. On adaptability: choice versus opportunity to be trustworthy

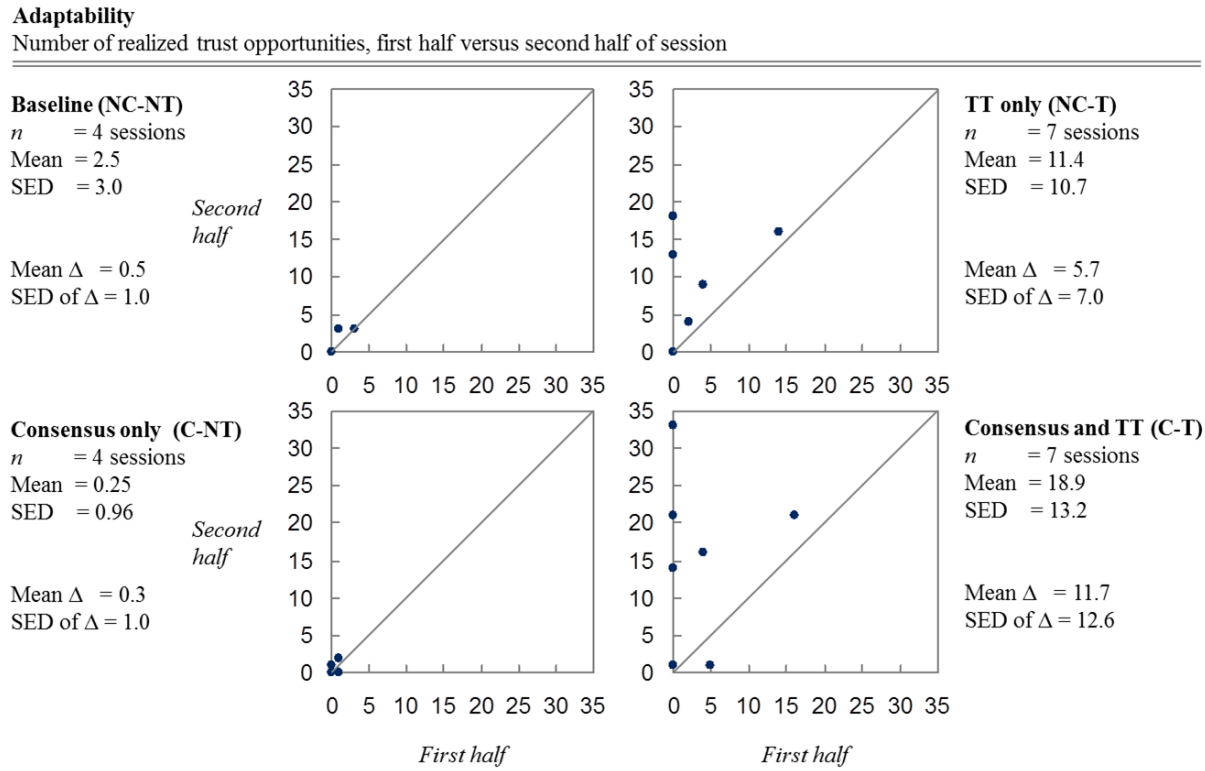
As discussed above, the naïve game theoretic approach proposes that turn-taking would not matter for adaptability. Requiring turn-taking should perform like the baseline, while requiring consensus and turn-taking should perform like requiring consensus only. The intuitive response is that creating more opportunities for reciprocity should foster more trustworthy behavior, and thus more trust. Counter-intuitively, however, a third of the trust opportunities in the "no turn-taking" treatments failed to pay off, not for lack of trustworthy choices by the counterpart, but because the counterpart never had an opportunity to reciprocate. Suppose the candidate who wins first is candidate "A". Then this "no opportunity" effect occurs (a) over three terms when the same candidate wins three times in a row ("AAA"), or (b) in the last term only when the candidates alternate ("ABA"). In effect, the three-term match expires and counterparts "retire" before they have the opportunity to reciprocate. Insofar as most trust between groups needs to be based on individual rather than group reputations, this imposes a substantial burden. The number of no-opportunity years on the "turn-taking" side, in contrast, is around 3%. This effect is missed entirely in models assuming infinitely-lived agents or parties (Dixit, Grossman and Gul 2000).

Our environment limits the size of the "no opportunity" effect by ensuring that leaders match up and run for three terms in a row. In the real world, however, it is commonplace for a new nominee to replace the losing candidate from the previous election. If this norm were always followed - as it has been in the United States since 1960 - then this effect would rule out this type of trust-based exchange.[16]

Figure 5 shows impressive first versus second half results for adaptability. In par-

[16]The last time a losing nominee won the nomination for a second run was Stevenson in 1956, to run against Eisenhower again after having lost to him in 1952.

Figure 5: Experimental results, focusing on adaptability



ticular, the two turn-taking treatments indicate significant and substantively large improvements in the number of realized trust opportunities, while the two non turn-taking treatments show no signs of change across the two halves of play.

D. On motivations: moving from bonding capital to bridging capital

While the propensity to trust in the lab leads some to propose that humans are more altruistic and less selfish than supposed, subsequent work has suggested that altruism is not unconditional and does not extend beyond the group (e.g. Boyd et al. 2003). Putnam (2000) distinguishes between bonding social capital, consisting of mutual trust between extended family, friends, and strangers who share a group identity, and bridging social capital, consisting of mutual trust between strangers across groups. While "bonding"

behaviors can mitigate collective action problems within groups, it can amplify conflict between them (Hardin 1997). Indeed, in other experiments, groups are less likely than individuals to engage in "bridging" behaviors and tend to be closer to the selfish predictions of rational choice in the way they treat one another, whether in trust games (Cox 2002; Kugler et al. 2007), dictator games (Cason and Mui 1997; Luhan and Sutter 2009), ultimatum games (Bornstein and Yaniv 1998), or centipede games (Bornstein, Kugler and Ziegelmeyer 2004).

Our results also find that the subjects are more "groupish" than rational choice would predict, but with the caveat that, with the right "bridging capital" in place, reciprocity between two officeholders can create trust between groups.

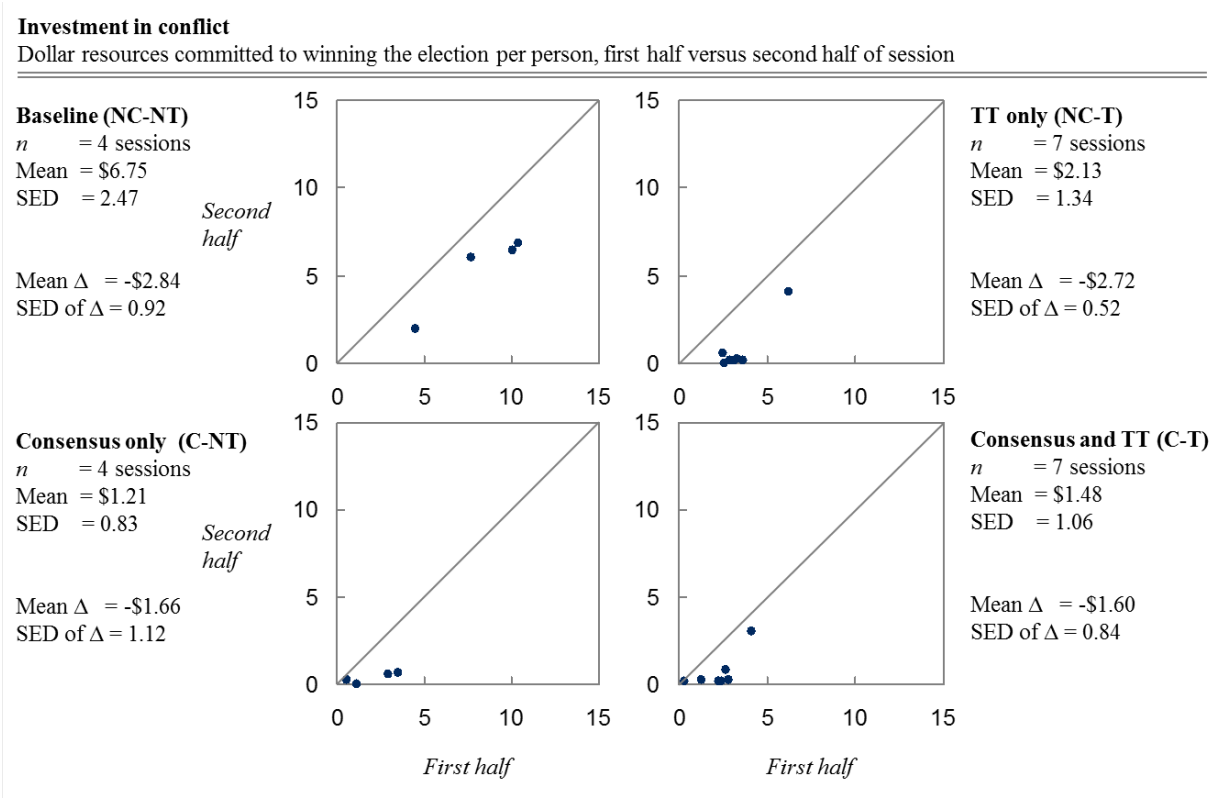
The selfish game-theoretic prediction is that subjects will invest the mixed strategy Nash equilibrium (MSNE) in the election, and then pick SPNE policies. The altruistic prediction is that subjects will invest nothing in the election, and then pick the most socially productive policy. A simple "groupish" model would propose that subjects over-invest relative to the MSNE up to the group optimum, and then pick SPNE policies.

In our experiment investments in conflict were greater than the individual MSNE. For the baseline (NC-NT), the investment per person was $6.75, more than 3 times the individual optimum of $2.16. For requiring consensus only (C-NT), the investment per person was $1.21, more than 1.5 times the individual optimum of $0.79.[17] This is common in other group conflict experiments featuring an interaction between intra-group and inter-group collective action problems (e.g. Goren and Bornstein 2000; Ahn, Isaac and Salmon 2011).

While levels of trust were much closer to the SPNE (e.g. zero trustworthiness and therefore zero trust) for the baseline (NC-NT) and the requiring consensus only (C-NT)

[17]For "requiring turn-taking only" (NC-T), the investment per person was $2.13, rather than the individual optimum of $0. Likewise, for "requiring consensus and turn-taking" (C-T), the investment per person was $1.48, rather than the individual optimum of $0.

Figure 6: Experimental results, focusing on investment in conflict



cases, requiring turn-taking only (NC-T) and requiring consensus and turn-taking (C-T) both saw substantial increases in trust above the relevant SPNE. In other words, behavior in the non-turn-taking environments is somewhere between selfish and groupish, as it is in the previous literature. *The turn-taking environments, however, provide enough "bridging capital" to overcome the distrust between groups.*

What form does that take and why does it work? Routine role reversal serves as as a kind of empathy-amplifier. Empathy occurs "when we suspend our single-minded focus of attention and instead adopt a double-minded focus of attention" to simulate the mental states of others (Baron-Cohen 2012, 15-16). The propensity for empathy not only varies by person, but by situation. For example, empathy tends to be greater when not in a hurry (Darley and Batson 1973), for familiar or immediately visible people (Iacoboni 2009), particularly those with whom we have reversed or could reverse roles (Haidt

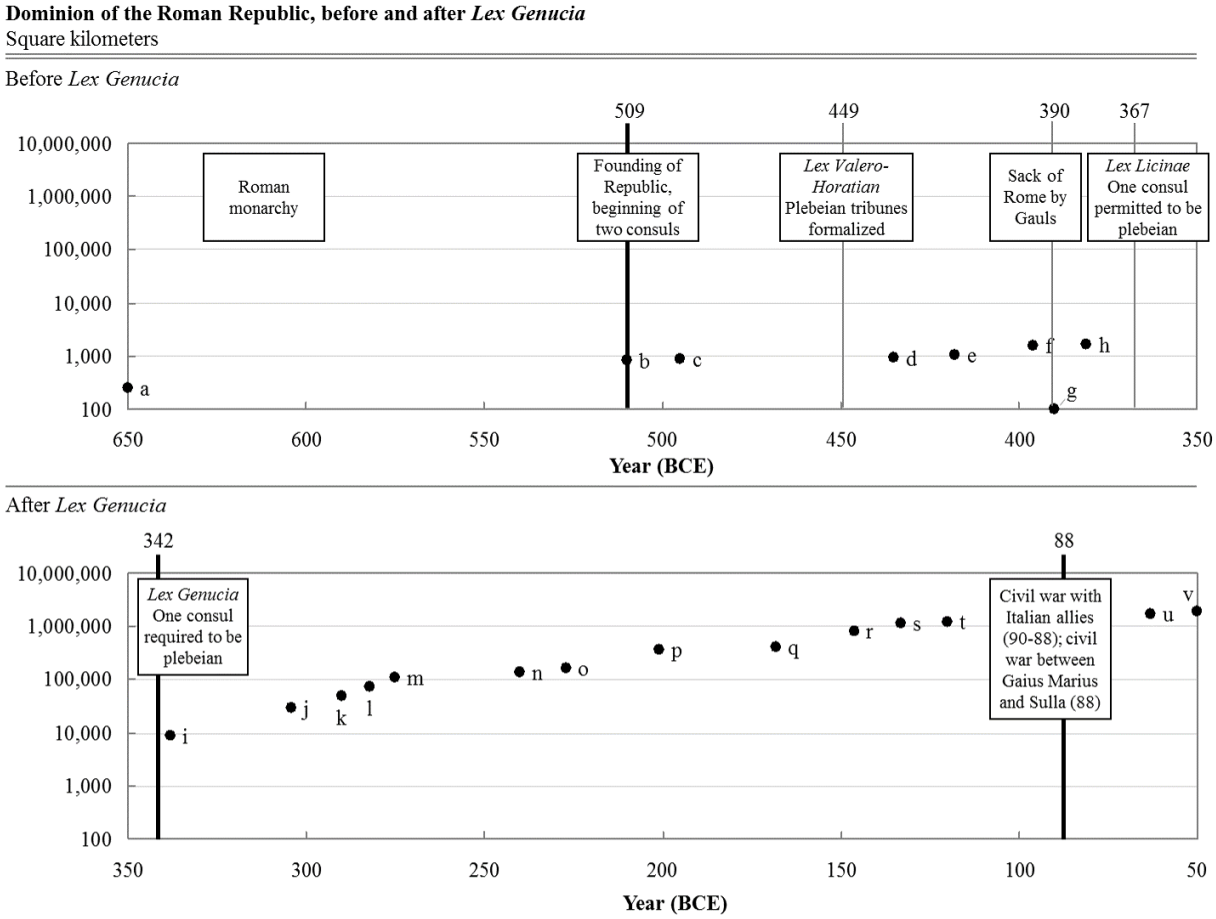
2006). Role reversal helps illuminate what risks and opportunities are common, rather than idiosyncratic. If on-going, it provides opportunities for mutual monitoring, reinforcement through positive reciprocity, and enforcement through negative reciprocity. Thus, mechanisms that cause "turn-taking" - where office-holders reverse roles multiple times at predictable intervals within the immediate future - can act as empathy-amplifiers, increasing how often and how deeply executives attempt to see their behavior from the perspective of an opposing counterpart (Lindskold 1978; Tyler 1993; Leach and Sabatier 2005).

Applications outside the lab: explaining the rise of Rome

The case of Rome is interesting for three reasons. First, within a 420 year time frame between approximately 510 and 90 BCE, the Romans experimented with both power-sharing and turn-taking institutions, with 170 years of power-sharing followed by 250 years of turn-taking. Second, the timing of the change from power-sharing to turn-taking in the middle of the 4th century BCE coincides with Rome's transformation from an ordinary city-state to something *extra*ordinary. Third, contemporary historians include the institutional change (from power-sharing to turn-taking) alongside other factors, but do not provide a causal account linking them. These accounts do not distinguish between causes of trust among opposing elites (e.g. the design of executive political institutions) and effects of that trust (e.g. on the design of other political institutions and the selection of public policies). On the causal side, these accounts do not distinguish between institutional structures that would tend to constrain trust, such as power-sharing, and those that would tend to amplify it further, such as turn-taking. By highlighting the impact of these distinctions, our experimental results help to shed new light on the rise of the Roman Republic, an ongoing debate for the last two millennium.

From the founding of the Republic around 509 BCE until the middle of the 4th cen-

Figure 7: The rise of Rome after *Lex Genucia*



From 100 to 1,000 km²:
a. Rome in 7th century (c. 650)
b. Rome at end of monarchy (c. 510)
c. Annexation of Ficulea, Crustumerium, and Nomentum (495)
d. Annexation of Fidanae (435)
e. Annexation of Labici (418)

From 1,000 to 10,000 km²:
f. End of Third Veientine War (396)
g. Sack of Rome by Gauls (390)
h. Annexation of Tusculum (381)
i. End of Latin War (338)
j. End of Second Samnite War (304)

From 10,000 to 100,000 km²:
k. End of Third Samnite War (290)
l. Battle of Populonia (282)
m. End of Pyrrhic War (275)

From 100,000 to 1,000,000 km²:
n. End of First Punic War (241)
o. End of Mercenary Wars (227)
p. End of Second Punic War (201)
q. End of Third Illyrian War (168)
r. End of Fourth Macedonian War (148), Achaean War (146) and Third Punic War (146)
s. Asia Minor annexed (133)
t. Gallia Transalpina annexed (120)

Above 1,000,000 km²:
u. End of Third Mithradic War (63)
v. End of Gallic Wars (50)

tury BCE, Rome was an ordinary city-state. Over this time, there was mounting conflict between patricians and plebeians, the so-called "Conflict of the Orders." Initially, plebeians used lynch mobs and informal strikes, or *secessio*, to defend their interests. In 449

BCE, the Leges Valeria-Horatiae put in place formal power-sharing institutions by recognizing plebeian offices. The most important of these were the tribunes of the people who could veto actions by the patrician consuls. Rome's dominion ebbed and flowed, but remained on the order of 1,000 square kilometers.[18] While Rome did not revert to a monarchy, internal strife seems to have prevented lasting military expansion. This is clearly shown in Figure 7, which plots the size of Roman dominion in square kilometers over 600 years, contrasting the 300 years before 350 BCE in the top panel and the 300 years after in the bottom panel.[19]

Beginning with the Latin War (341-338 BCE), Rome begins to win and maintain control of territory at an exponential rate. Over the next 200 years, there is a three order-of-magnitude increase in scale: from the city-state of Rome (1,000 square kilometers) to the dominant power of central Italy (10,000 square kilometers) by 338 BCE, then all of the Italian peninsula (100,000 square kilometers) by 275 BCE, and then much of the Mediterranean (1,000,000 square kilometers) by 146 BCE. Contemporary historians (e.g. Cornell 1995; Forsythe 2005; Beard 2015) point to the middle of the fourth century as the turning point for the Roman Republic, focusing on the thirty years between 367 BCE (when Lex Licinio-Sexta began the settlement between the patricians and plebeians) and 338 BCE (when Rome won the Latin War and arrived at a settlement that set the precedent for giving conquered elites a standard set of obligations, rights, and liberties).

In addition to the consensus about *when* Rome began its rise, there is a consensus about *why*. Commonly cited proximate causes of Roman success - with minor variations - correspond to the three dependent variables we have emphasized above:

[18]The positive deviation would be the capture of Veii in 396 BCE which would have increased area up to 1600 square kilometers. The negative deviation was the sack of Rome by the Gauls in 390 BCE (Cornell 1995).

[19]Data drawn from Cornell (1995) and Taagepera (1978).

1. *Agility to overcome acute challenges*, especially the leadership capabilities to mobilize and manage resources and logistics at a previously unknown scale.[20]

2. *Trust to overcome chronic challenges*, especially the settlement of conflict between patricians and plebeian elites. This trust was built by requiring plebeian access to magistracies;[21] making the membership of the Senate the plebeian and patrician alumni of those magistracies, rather than ad hoc advisers chosen by consuls; creating a standard career progression for ambitious plebeians or patricians; opening up some plebeian access to public lands; and abolishing debt slavery.[22]

3. *Adaptability to overcome constitutive challenges*, particularly the ability to retain and leverage the loyalty of conquered elites with a relatively standard set of obligations, rights, and liberties, including both the obligation to put troops in the field in future campaigns and the right to share in the spoils.[23]

Why did these three factors come together in the middle of the 4th century BCE?

[20]For example, organizing 40,000 Romans and allies for the Battle of Sentinum in 295 BCE. Beard (2015) notes this is a different order of magnitude than skirmishes from the previous century, and that this scale requires the "complex and demanding logistics of equipment, supply, and animal transport." Beard estimates that a force this size would have required 50 tonnes of wheat per day. The extraordinary scale of leadership and organizational discipline is also noted in Cornell (1995) and Forsythe (2005). The discussion of this factor goes back to Livy's discussion of the number of capable Roman leaders at a given time and across generations, in contrast to Alexander the Great, a single, charismatic leader in one generation (Beard 2015).

[21]Consuls in 342 BCE, censors in 339 BCE, and pontifices and augurs in 300 BCE.

[22]This is discussed across Cornell (1995), Forsythe (2005), and Beard (2015). The substrate of the conflict was differences in political rights, differences in access to public lands, and the threat of debt slavery (Cornell 1995). Classical historians (e.g. Polybius and Livy) made the observation that overcoming the Conflict of the Orders and achieving a new level of stability and order among elites was key to the rise of Rome.

[23]This principle began with the settlement of the Latin Wars in 338 BCE and applied going forward to conquered cities or newly founded colonies. Again, this is discussed across Cornell (1995), Forsythe (2005), and Beard (2015). Arnold Toynbee noted that the innovations of this settlement "gave the Roman commonwealth the maximum capacity for expansion, combined with the maximum solidity of structure" (Toynbee 1965).

Though there is a consensus on these *proximate* causes of the rise of Rome, there is little discussion on the *ultimate* causes. While we are constrained by the incompleteness and ambiguity of the historical record, theory and experiments on trust and trustworthiness generally - and the time horizon trilemma specifically - allow us to glean more insight from what we already know.

One general insight comes from the theory and empirical work in North, Wallis and Weingast (2009) and Acemoglu and Johnson (2005): it is easier for trust to take root among a few powerful elites based on personal relationships, to come to consensus on more impersonal rules over time to regulate their interactions with one another, and then to use the impersonal or abstract nature of the rules to include or exclude other counterparties as needed. In the case of the Roman Republic, our hypothesis is (1) that *sine qua non* was requiring one consul to be plebeian in 342 BCE, as this fostered trust between one patrician and one plebeian leader at the highest level of the state; (2) that requiring collegiality or pairing between patricians and plebeians for other magistracies further increased bridging capital by broadening the set of patrician-plebeian partnerships; (3) that shared access to the same magistracies enabled increasingly standard career paths that made it possible to build leadership capabilities at scale; (4) that the broader set of patrician-plebeian partnerships made it easier to build elite consensus around the rights, liberties, and obligations of all citizens, rather than officeholders only; and (5) that this consensus created a stable tradition of public law that - as an unintended consequence - could be extended to conquered elites.

On the first point, it is striking that the shift from a power-sharing approach to a turn-taking approach, beginning with Lex Licinio-Sexta in 367 BCE and culminating in Lex Genucia in 342 BCE, immediately precedes the rise of the Republic. This is consistent with the experimental results shared here: requiring consensus can constrain trust but requiring turn-taking can amplify it. We should expect different results from turn-taking

within a particular office than from power-sharing across offices.

On the second point, Forsythe (2005) finds many cases after 367 BCE where a particular patrician-plebeian pair returns to office at the same time - far too many to be coincidental - suggesting pairing as a "ticket": in the 100 years after 367 BCE, 10 patrician-plebeian pairs held the consulate two or more times together. An additional 5 pairs held the consulate and the censorship together, while an additional 5 pairs *of families* held the consulate two or more times together, and so forth.

On the third point, the standardization of offices made possible the *cursus honorum*, or "course of offices," that channeled ambition within a standard set of leadership experiences, which were typically re-allocated on an annual basis. [24]

On the fourth point, after *Lex Ovinia* any patrician or plebeian who held the rank of praetor or above would win lifetime membership in the Senate, meaning that the patrician-plebeian balance in the magistracies would be further reflected in the Senate. This made the division between patrician and plebeian matter less and less; in its place, there was a new class of citizens defined by shared experiences in leadership roles, the new *nobility*.[25]

On the fifth point, it is notable how *stable* the tradition of obligations, rights, and liberties became. From 342 BCE to 89 BCE, as the Roman Republic grew by three orders of magnitude, more than 400 unique patrician and plebeian executives would take their turn in the consul office, and then peacefully surrender it.[26] The stability at the core of

[24]This includes the military tribune, quaestor, aedile, praetor, consul, and censor. The third volume of Broughton (1986) provides a catalog of "careers" of hundreds of Roman leaders.

[25]The exact date of *Lex Ovinia* is unknown. Cornell (1995) notes that it would fall after 339 BCE, when Lex Publilia required that one censor be plebeian, and before 318 BCE, when we know the Senate was selected by censors.

[26]As a point of comparison, the longest run of peaceful transfers of power in the United States has been the 150 years since the Civil War, and included 27 unique presidents; in the UK it has been since the Revolution Settlement, and included 53 unique prime ministers, starting with Walpole in 1721. As of 2008, over 60 countries in the world - including Russia and China - had

Roman public law made it possible to innovate on the margins, extending *civitas sine suffragio* to conquered elites or colonies.

Conclusion

This article makes three contributions to the literature on institutional design, trust and bargaining among opposing elites, and experimental political science.

First, we conceptualize the time horizon trilemma, complicating theories of inter-elite cooperation by adding a set of countervailing challenges to the well-known credible commitment problem. Acknowledgment of these challenges helps explain stubborn bargaining failures seen empirically, and challenges assumptions built into commonly cited rational choice models and core claims in the literature on institutional design. Our results show the promise of mechanisms inspired by the Roman consulate to push elite behavior towards inter-group cooperation while avoiding the pitfalls of brinkmanship and loss of agility that accompany power-sharing.

Second, we innovate in the experimental lab: very few existing experimental studies pair election with policy-making games to assess how electoral institutions affect the shape of policy and, in particular, trust and trustworthiness.[27] While a few studies have assessed new electoral options in the lab for proof of concept (e.g. Casella and Palfrey 2006), these do not seek to test prominent alternatives against one another, hampering our ability to understand trade-offs among them. Our design serves as a stepping stone between more basic experiments (e.g. trust games, ultimatum games, Blotto contests) and more complex ones (e.g. capturing the nuances of elite conflict and trust-formation). Additional treatments could create a path from more basic research to more applied re-

not had a party lose an election and peacefully transfer power as a result (Przeworski 2015).

[27]One exception is Rodet (2015).

search. To simplify the environment further, we could replace the three-battlefield Blotto environment with a one-battlefield case, a lottery, or a coin toss. To complicate the environment further, we could enable retrospective voting, making leaders accountable to elites or mass segments; add multiple issues to require dynamic prioritization, so that agility matters more; and give incumbents the ability to prevent elections and/or allow the opposition to instigate violence.

Third, we provide new perspective on a 2,100 year-old debate regarding the ultimate causes of the rise of the Roman Republic. It was a puzzle to historians in the second and first centuries BCE, and - while there is now much more consensus on the timing and proximate causes of Rome's rise - it remains a puzzle today. The historical record is too incomplete to know for sure, but the results here suggest 342 BCE may have been the turning point for the rise of the Roman Republic, and that its rise can be traced to institutional changes involving rotation in office.

Our research on turn-taking, however, would lack broader impact if there were no practical prospect of "field testing" the mechanism in the real world. There are three encouraging trends: (1) turn-taking builds on developments in the private sector, (2) it could be appealing to term-limited mayors or governors, and (3) there is dissatisfaction with existing solutions at the national level and a willingness to try appropriately vetted ideas. We address each in turn.

First, there has been ample experimentation with the idea of "pairing" in the private sector to improve the efficiency and effectiveness of behavior, including among surgeons (Jari and Shelbourne 2002), computer programmers (Williams and Kessler 2002), managers (Canner 2008), and CEOs (Arena, Ferris and Unlu 2011). World class organizations like the Mayo Clinic and Toyota have used paired leadership across hundreds of roles over the course of decades.[28] Initial field tests of the ideas presented here may take place

[28]The Mayo Clinic has used paired leadership between physicians and administrators at the de-

in the private sector, where the target population is larger and interested parties could opt in without widespread and contentious political changes.

Second, term-limited mayors or governors could opt into a pilot, both for public-spirited reasons and careerist reasons (e.g. replacing term limits with turn-taking, thereby extending their tenure in office). It is not unusual for executives to look into revising the term limits that bind them, as Michael Bloomberg did in New York. Half of the mayors of the largest 100 cities in the United States are term limited, and - as most are limited to two terms - roughly 25 of the 50 tend to be in their last term at any given time. These 25 - as well as hundreds of their term-limited peers in smaller cities - would be the target audience for pilot tests. While it is commonplace for management ideas developed in the private sector to influence the public sector after a few years (Christensen et al. 2007), interest on the part of incumbent executives would accelerate the process.

Third, and finally, there is dissatisfaction with existing solutions at the national level and a willingness to try appropriately vetted ideas. Scholars and policymakers engage in robust debates about the relative performance of alternative democratic institutions, particularly in polarized or divided societies. Some evaluate or recommend variations on single member districts that use a different aggregation rule - such as instant run-off, approval voting, or range voting - or variations on multi-member districts that use one of many proportional representation rules (e.g. Guinier 1994; Lijphart 2004). At the same time, political exigencies have led to "constitutional moments" in places a varied as Iraq, Kenya, Zimbabwe, Egypt, Sudan, and Afghanistan, among many others. If the proposed ideas have been field-tested in the lab, in the private sector, and at local levels

partment level, campus level and enterprise level since the 1950s, modeled after the partnership between Will Mayo and Harry Harwick. Toyota has used paired leadership between Japanese and native managers at the functional manager, general manager, and VP levels in 40+ plants outside Japan, mostly since the 1980s. This is based on author interviews with existing managers at the Mayo Clinic and Toyota.

of government, they could influence these national constitutional moments for the better.

There is an appropriately long path to impact at the national scale, but one that offers many opportunities for discovery, refinement, and impact at smaller scales. We have already taken a step down the path: this study tested the initial plausibility of the turn-taking mechanism and confirmed that turn-taking increased trust, conserved agility, and increased adaptability. As the Romans may have realized, cooperation is more likely to emerge when opposing leaders have many opportunities to trust and be trustworthy in turn.

References

Acemoglu, Daron and Simon Johnson. 2005. "Unbundling Institutions." *Journal of Political Economy* 113(5):949–995.

Ahn, T.K., R. Mark Isaac and Timothy C. Salmon. 2011. "Rent seeking in groups." *International Journal of Industrial Organization* 29(1):116 – 125.

Arena, Matteo P., Stephen P. Ferris and Emre Unlu. 2011. "It Takes Two: The Incidence and Effectiveness of Co-CEOs." *The Financial Review* 46:385–412.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Baron-Cohen, Simon. 2012. *The science of evil: On empathy and the origins of cruelty*. Basic books.

Bates, Robert. 2008. *When Things Fell Apart. State Failure in Late-Century Africa.* New York: Cambridge University Press.

Beard, Mary. 2015. *SPQR: A History of Ancient Rome*. Liveright Publishing Corporation.

Bednar, Jenna. 2008. *The Robust Federation*. Cambridge: Cambridge University Press.

Berg, Joyce, John Dickhaut and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10:122–142.

Borel, Emile. 1921. "La théorie du jeu et les équations intégrales à noyau symétrique." *Comptes Rendus de l'Académie des Sciences* 173(1304-1308):58.

Bornstein, Gary and Ilan Yaniv. 1998. "Individual and group behavior in the ultimatum game: Are groups more 'rational' players?" *Experimental Economics* 1(1):101–108.

Bornstein, Gary, Tamar Kugler and Anthony Ziegelmeyer. 2004. "Individual and group decisions in the centipede game: Are groups more rational players?" *Journal of Experimental Social Psychology* 40(5):599–605.

Boyd, Robert, Herbert Gintis, Samuel Bowles and Peter J Richerson. 2003. "The evolution of altruistic punishment." *Proceedings of the National Academy of Sciences* 100(6):3531–3535.

Broughton, T Robert S. 1986. *Magistrates of the Roman Republic*. Vol. 3: Supplement American Philological Association.

Bueno de Mesquita, Bruce and Alastair Smith. 2010. "Leader Survival, Revolutions, and the Nature of Government Finance." *American Journal of Political Science* 54:936–950.

Cachon, Gérard and Christian Terwiesch. 2009. *Matching supply with demand*. Vol. 2 McGraw-Hill Singapore.

Canner, Niko. 2008. "The Issue: For Cognizant, Two's Company.".

Casella, Alessandra, Andrew Gelman and Thomas R. Palfrey. 2006. "An Experimental Study of Storable Votes." *Games and Economic Behavior* 123-154.

Cason, Timothy N and Vai-Lam Mui. 1997. "A Laboratory Study of Group Polarisation in the Dictator Game." *The Economic Journal* 107(444):1465–1483.

Cherry, Todd L., Peter Frykblom and Jason F. Shogren. 2002. "Hardnose the Dictator." *American Economic Review* 92(4):1218–1221.

Christensen, Tom, Per Lægreid, Paul G Roness and Kjell Arne Røvik. 2007. *Organization theory and the public sector: instrument, culture and myth*. Routledge.

Cornell, Tim. 1995. *The beginnings of Rome: Italy and Rome from the Bronze Age to the Punic Wars (c. 1000-264 BC)*. Psychology Press.

Cowen, Tyler. 2005. "Self-Deception as the Root of Political Failure." *Public Choice* 124:437–51.

Cox, James C. 2002. Trust, reciprocity, and other-regarding preferences: Groups vs. individuals and males vs. females. In *Experimental Business Research*. Springer pp. 331–350.

Darley, John M and C Daniel Batson. 1973. "'From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior." *Journal of Personality and Social Psychology* 27(1):100.

Dixit, Avinash, Gene M Grossman and Faruk Gul. 2000. "The Dynamics of Political Compromise." *Journal of Political Economy* 108(3):531–568.

Fearon, James D. 1995. "Rationalist explanations for war." *International Organization* 49(3):379–414.

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10(2):171–178.

Flores, Thomas Edward and Irfan Nooruddin. 2012. "Electing peace: The effect of elections on post-conflict reconstruction." *Journal of Politics* 74(2):558–570.

Forsythe, Gary. 2005. *A critical history of early Rome: from prehistory to the first Punic War*. Univ of California Press.

Garfinkel, Michelle R and Stergios Skaperdas. 2007. "Economics of conflict: An overview." *Handbook of defense economics* 2:649–709.

Glaeser, Edward L, David I Laibson, Jose A Scheinkman and Christine L Soutter. 2000. "Measuring trust." *Quarterly Journal of Economics* pp. 811–846.

Golman, Russell and Scott E. Page. 2009. "General Blotto: Games of Allocative Strategic Mismatch." *Public Choice* 138:279–299.

Goodin, Robert E. 1996. "Institutionalizing the public interest: The defense of deadlock and beyond." *American Political Science Review* pp. 331–343.

Goren, Harel and Gary Bornstein. 2000. "The Effects of Intragroup Communication on Intergroup Cooperation in the Repeated Intergroup Prisoner's Dilemma (IPD) Game." *Journal of Conflict Resolution* 44(5):700–719.

Guinier, Lani. 1994. *The tyranny of the majority: Fundamental fairness in representative democracy*. Free Press New York.

Haidt, Jonathan. 2006. *The happiness hypothesis: Finding modern truth in ancient wisdom*. Basic Books.

Haidt, Jonathan. 2013. *The righteous mind: Why good people are divided by politics and religion*. Random House LLC.

Hardin, Russell. 1997. *One for All: The Logic of Group Conflict*. Princeton University Press.

Hetherington, Marc J and Jason A Husser. 2012. "How trust matters: The changing political relevance of political trust." *American Journal of Political Science* 56(2):312–325.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat and Vernon Smith. 1994. "Preferences, Property Rights, and Anonymity in Bargaining Games." *Games and Economic Behavior* 7(3):346–380.

Iacoboni, Marco. 2009. "Imitation, empathy, and mirror neurons." *Annual review of psychology* 60:653–670.

Jari, Sanjiv and K Donald Shelbourne. 2002. "Simultaneous bilateral anterior cruciate ligament reconstruction." *The American journal of sports medicine* 30(6):891–895.

Keele, Luke. 2007. "Social capital and the dynamics of trust in government." *American Journal of Political Science* 51(2):241–254.

Kovenock, Dan and Brian Roberson. 2012. "Coalitional Colonel Blotto Games with Application to the Economics of Alliances." *Journal of Public Economic Theory* 14(4):653–676.

Kugler, Tamar, Gary Bornstein, Martin G Kocher and Matthias Sutter. 2007. "Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy." *Journal of Economic psychology* 28(6):646–657.

Leach, William D. and Paul A. Sabatier. 2005. "To Trust an Adversary: Integrating Rational and Psychological Models of Collaborative Policymaking." *American Political Science Review* 99(4):491–503.

Lijphart, Arend. 2004. "Constitutional Design for Divided Societies." *Journal of Democracy* 51(2):96–109.

Lindskold, Svenn. 1978. "Trust development, the GRIT proposal, and the effects of conciliatory acts on conflict and cooperation." *Psychological Bulletin* 85(4):772.

List, John A. 2007. "On the Interpretation of Giving in Dictator Games." *Journal of Political Economy* 115(3):482–493.

Logan, Gordon D. 1992. "Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18(5):883.

Luhan, Wolfgang J., Martin G. Kocher and Matthias Sutter. 2009. "Group polarization in the team dictator game reconsidered." *Experimental Economics* pp. 26–41.

McCabe, Kevin A, Mary L Rigdon and Vernon L Smith. 2003. "Positive reciprocity and intentions in trust games." *Journal of Economic Behavior & Organization* 52(2):267–275.

Merolla, Jennifer, Michael Munger and Michael Tofias. 2005. "In Play: A Commentary on Strategies in the 2004 U.S. Presidential Election." *Public Choice* 123(1-2):19–37.

Moe, Terry M. and William G. Howell. 1999. "Unilateral Action and Presidential Power: A Theory." *Presidential Studies Quarterly* 29(4):850–873.

North, Douglass, John Wallis and Barry Weingast. 2009. *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge University Press.

Ostrom, Elinor. 2005. *Understanding Institutional Diversity*. Princeton University Press.

Ostrom, Elinor and James Walker, eds. 2005. *Trust and Reciprocity: Interdisciplinary Lessons for Experimental Research*. Russell Sage Foundation.

Page, Scott E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press.

Pildes, Richard H. 2011. "Why the Center Does Not Hold: the causes of hyperpolarized democracy in America." *Cal. L. Rev.* 99:273.

Posner, Eric and Adriana Vermeule. 2010. *The Executive Unbound: After the Madisonian Republic*. Oxford University Press.

Przeworski, Adam. 2015. "Acquiring the Habit of Changing Governments Through Elections." *Comparative Political Studies* 48(1):101–129.

Putnam, Robert D. 2000. *Bowling alone: The collapse and revival of American community*. Simon and Schuster.

Robinson, James A. 1998. "Theories of "Bad Policy"." *Policy Reform* 2(1):1–46.

Rodet, Cortney S. 2015. "An experiment in political trust." *Journal of Behavioral and Experimental Economics* 57:17–25.

Rodrik, Dani, Arvind Subramanian and Francesco Trebbi. 2004. "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development." *Journal of Economic Growth* 9(2):131–165.

Rosenthal, Robert W. 1981. "Games of perfect information, predatory pricing and the chain-store paradox." *Journal of Economic theory* 25(1):92–100.

Slonim, Robert and Alvin E Roth. 1998. "Learning in high stakes ultimatum games: An experiment in the Slovak Republic." *ECONOMETRICA-EVANSTON ILL-* 66:569–596.

Sunstein, Cass R. forthcoming. "Partyism." *University of Chicago Legal Forum* .

Szentes, Balázs and Robert W. Rosenthal. 2003. "Three-object two-bidder simultaneous auctions: chopsticks and tetrahedra." *Games and Economic Behavior* 44(1):114 – 133.

Taagepera, Rein. 1978. "Size and duration of empires: systematics of size." *Social Science Research* 7(2):108–127.

Toynbee, Arnold Joseph. 1965. *Hannibal's legacy*. Vol. I Oxford University Press.

Tsebelis, George. 2002. *Veto Players – How Political Institutions Work*. Princeton: Princeton University Press.

Tyler, Tom. 1993. *Social Psychology in Organizations: Advances in Theory and Practice*. Prentice-Hall chapter The social psychology of authority.

Walter, Barbara F. 2002. *Committing to Peace*. Princeton University Press.

Weingast, Barry R. 1997. "The Political Foundations of Democracy and the Rule of Law." *American Political Science Review* 91(2):245–263.

Williams, Laurie and Robert Kessler. 2002. *Pair programming illuminated*. Addison-Wesley Longman Publishing Co., Inc.

Appendix 1

Welcome!

Thanks for completing the pre-play survey.

Please turn off all electronic devices and place them in your bags or under your desk. Please do not talk during the experiment. If you have any questions, raise your hand and the experimenter will come to personally assist you.

Today's experiment involves several tasks. All participants will receive a payment of $5 for showing up on time and completing all of the tasks. Participants can win a bonus based on performance in the games we will play, up to ≈ $30 more. At the end of the study, you will be paid privately in cash.

1. Overview

For today's session, you will be randomly assigned to a team (either Red or Blue) with three other participants. This team will play 8 games with an opposing team. The teams will be fixed for the whole session today.

Each game consists of three terms, and each term includes both an election and four years of policy-making (see Figure 1).

For each game, each team will have one player assigned as the leader. The leader will be responsible for making policy decisions during each of the 3 terms per game. Everyone will have 2 opportunities to be the leader throughout the course of the session, once in the first half and once in the second half, in an order that is randomly assigned.

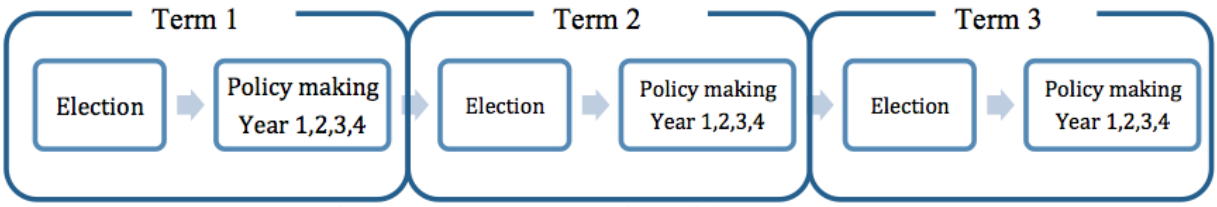


Figure 1. One game consists of three terms.

To determine your bonus, we will pick one of the games you played at random, and pay you for that game only. So do your best in every game!

2. Instructions for each game

a. Endowments and payoffs. You will start each term with an endowment of $10. Your payoff at the end of the term will be: the resources you started with **minus** any resources you used up in the election **plus** any resources you won from policy payoffs.

b. Rules for each electoral contest. There are three "battlefields" in the electoral contest. Whoever wins the most battles wins control of the office for the term. Whoever commits the most resources to a battlefield wins the battle. Ties are settled by a coin toss.

Any resources you commit in the election get used up, whether you win or lose. You will have 40 seconds to make your commitments final. At the halfway point (i.e. 20 seconds in, with 20 seconds to go), you will see a snapshot of your opponent's choice so far, and they will see a snapshot of your choice. This gives you 20 more seconds to react.

c. Rules for policy-making

(For requiring turn-taking versions)

Whoever wins the election will be "in" for the first and third years of the term, but roles will be reversed for the second and fourth years.

(For not requiring turn-taking versions)

Whoever wins the election will be "in" for all four years of the term.

(For requiring consensus versions)

The "in" leader proposes policy payoffs for the year to the "out" leader. The "out" leader can either accept or reject the offer. If the "out" leader rejects the offer, neither side receives a policy payoff for the year. If the "out" leader accepts the offer, both sides receive the agreed-upon policy payoffs.

(For no requiring consensus versions)

The "in" leader chooses policy payoffs for the year.

(For all versions)

The possible policy payoffs are:

Payoff for own side ("Ins")	Payoff for other side ("Outs")
$5	$0
$4	$2
$3	$4
$2	$6
$1	$8
$0	$10

Suppose the second policy is chosen for the year. Then each member of the "in" team, including the leader, would get $4 and each member of the "out" team would get $2 for that year.

3. Payment

Your payoff at the end of each game will be the **average of your payoffs for each term.**

Remember, we will pick one of the games you play at random, and pay you for that game only. So do your best in every game!

For example, suppose we randomly choose the sixth game to set your payoff at the end of the session today, and in that game you ended up with $25, $20, and $15 in terms 1, 2, and 3. Your payoff for the game is $20. So you would walk away with that $20 bonus plus the $5 payment for showing up.

To make sure you understand these instructions, we will give you a 1 minute comprehension quiz and let you play two practice terms. After that, the rest of the experiment will begin.

GOOD LUCK!

Appendix 2

This appendix discusses the process for deriving numerical solutions for the election game given the stakes of the policy-making games. The game featured in the design is a non-constant-sum or "all-pay" Colonel Blotto game with three battlefields between two groups each of four players, using dollar unit bids. There a no known closed form solution for this game, or immediately adjacent variations. We have to reduce the number of players down to two and make the dollar bids continuous in order to reach a variant of the game with a closed form analytical solution (Szentes and Rosenthal 2003; Kovenock and Roberson 2012). This is one step removed from the original constant-sum Colonel Blotto game introduced by Borel (1921). Here we explain incremental changes to the MSNE as it moves from (a) the original two-player three-battlefield Blotto game to (b) the non-constant sum version to (c) the two-player all dollar units version to (d) the eight-player version numerically to treat the variant we use in the lab.

The original Colonel Blotto game was constant-sum in that resources were "use it or lose it": resources that are not allocated to one of the battles are lost. For a three battlefield Blotto game between two players, the unique symmetric MSNE is to randomly allocate all resources across all three battlefields using a uniform distribution.

The non-constant sum version resembles an all-pay auction or a Tullock lottery in that resources committed to winning have an opportunity cost; they could simply be retained. The 3-battlefield version of Colonel Blotto is isomorphic to what Szentes and Rosenthal (2003) call a simultaneous "pure chopstick" auction, where chopsticks are suggestive of identical objects that are useless except in pairs. Winning one battlefield alone is worth nothing, winning two is worth the full value of the prize, and winning three is worth nothing incremental to two. Here the unique symmetric MSNE is for both sides to randomly choose a budget using the uniform distribution between 0 and the value of the

prize, $v_{1st} - v_{2nd}$, and then to randomly allocate it across the three battlefields, subject to the constraint that the battlefield with the most resources has no more than the sum of the lesser two.

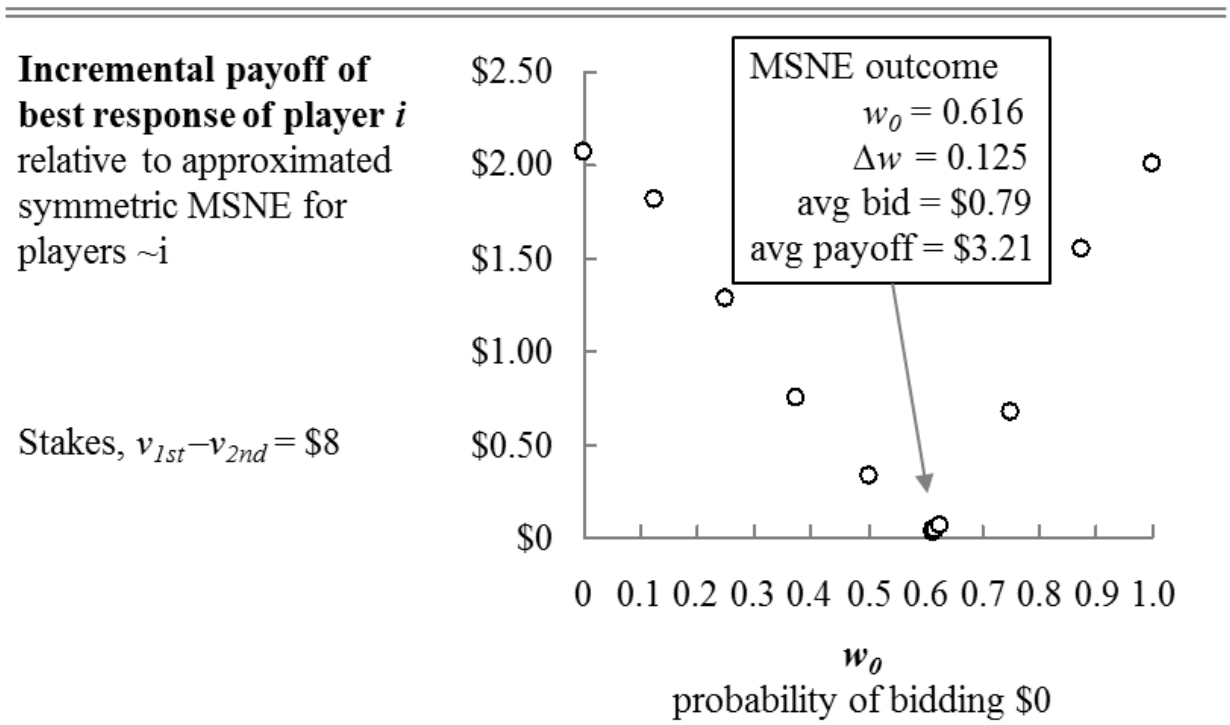
The intuition on the budget is that each seeks to make the other indifferent over the full range from 0 to the stakes of the prize. To make the expected value of all these bids equal, the benefit from the increased probability of winning must be equal to the loss from increasing one's bid. This means the uniform distribution. The intuition on the concentration is that this is the point at which the increased probability of dominating one battlefield is outweighed by the increased probabilities of losing the other two. Otherwise, one is spending more to win one particular battlefield (which alone is worth nothing) rather than on any particular two (which are needed to win the war).

This solution is only incrementally different once we move from continuous to dollar units. When neither side is willing to pay the full value of the prize, and each seeks to make the other indifferent over the full range, from 0 to the stakes of the prize minus one dollar, the result is that the average bid should be \$0.50 less than it would be in the continuous environment.

The eight-player version introduces a wedge between the optimal budget for the group to bid and the optimal budget for the individual to bid. The optimal budget for the group is the same as it is in the two-player game, but the individual has incentives to free-ride on his fellow group members. To numerically derive the symmetric mixed strategy Nash equilibrium, we search for the probability weight assigned to playing zero, w_0, constrained by the fact that the incremental probability of committing each dollar unit up to the maximum bid, x_{max}, must be $1/v$, that makes player i indifferent over the same range. The maximum bid, x_{max}, is $\|\frac{1-w_0}{\Delta p}\|$, and every positive commitment less than x_{max} is played with probability Δw, and x_{max} itself is played with the probability $1 - w_0 - (x_{max} - 1) * (\Delta w)$.

The example when $v_{1st} - v_{2nd} = 8$ is displayed in Figure A1. The incremental payoff of the best response of player i relative to the symmetric strategy of players $\approx i$ is approximately zero when w_0 = 0.616, and therefore w_1 = 0.125, w_2 = 0.125, w_3 = 0.125, and w_4 = 0.009. The weighted average of these is \$0.79. By contrast, the individual contribution that is optimal for the group would be to set w_0 to 0.128, implying an average contribution of \$3.48. The overall process for determining optimal budget distributions is visualized below in Figure A2.

Figure A1: Mixed Strategy Nash Equilibrium Outcome



(1) For a given budget, groups choose among all possible distributions across 3 battlefields *Example*

Probability of winning election for group 1
as a function of the distribution of resources across battlefields

$11 budget chosen by group 2			10	9	8	8	7	7	6	6	6	5	5	5	4	4
			0	1	2	1	3	2	4	3	2	5	4	3	4	3
			0	0	0	1	0	1	0	1	2	0	1	2	2	3
11	0	0	0.42	0.67	0.67	1.00	0.67	1.00	0.67	1.00	1.00	0.67	1.00	1.00	1.00	1.00
10	1	0	0.25	0.38	0.50	0.67	0.50	0.83	0.50	0.83	1.00	0.50	0.83	1.00	1.00	1.00
9	2	0	0.33	0.38	0.38	0.33	0.50	0.50	0.50	0.67	0.67	0.50	0.67	0.83	0.83	1.00
9	1	1	0.00	0.25	0.33	0.42	0.33	0.67	0.33	0.67	1.00	0.33	0.67	1.00	1.00	1.00
8	3	0	0.33	0.50	0.38	0.50	0.38	0.33	0.50	0.50	0.33	0.50	0.67	0.50	0.67	0.67
8	2	1	0.00	0.17	0.38	0.25	0.33	0.38	0.33	0.50	0.67	0.33	0.50	0.83	0.83	1.00
7	4	0	0.33	0.50	0.50	0.67	0.38	0.50	0.38	0.33	0.33	0.50	0.50	0.33	0.33	0.17
7	3	1	0.00	0.17	0.33	0.33	0.38	0.38	0.33	0.38	0.33	0.33	0.50	0.50	0.67	0.67
7	2	2	0.00	0.00	0.33	0.00	0.50	0.25	0.33	0.33	0.42	0.33	0.33	0.67	0.67	1.00
6	5	0	0.33	0.50	0.50	0.67	0.50	0.67	0.38	0.50	0.50	0.25	0.17	0.17	0.00	0.00
6	4	1	0.00	0.17	0.33	0.33	0.33	0.50	0.38	0.38	0.50	0.33	0.38	0.33	0.33	0.17
6	3	2	0.00	0.00	0.17	0.00	0.50	0.17	0.50	0.38	0.25	0.33	0.33	0.38	0.50	0.67
5	5	1	0.00	0.17	0.33	0.33	0.33	0.50	0.33	0.50	0.67	0.42	0.25	0.33	0.00	0.00
5	4	2	0.00	0.00	0.17	0.00	0.33	0.17	0.50	0.33	0.33	0.67	0.38	0.38	0.25	0.17
5	3	3	0.00	0.00	0.00	0.00	0.33	0.00	0.67	0.33	0.00	0.67	0.50	0.25	0.33	0.42
4	4	3	0.00	0.00	0.00	0.00	0.17	0.00	0.67	0.17	0.00	1.00	0.67	0.17	0.42	0.17

$10 budget chosen by group 1

(2) Within those distributions, the "balanced" distributions (where resources on the largest are equal or less than the sum of those on the smaller two) dominate the "non-balanced"

Probability of winning election for group 1
as a function of balanced or non-balanced distribution

Group 2	Non-balanced	Balanced
Non-balanced	0.43	0.69
Balanced	0.12	0.38

Group 1

(3) While there is some randomness when groups' chosen budgets are near equal in size, the greater the difference the more likely the larger can win with certainty

Probability of winning election for group 1
as a function of budgets randomly chosen distributions across battlefields
constrained so that largest is not greater than sum of other two

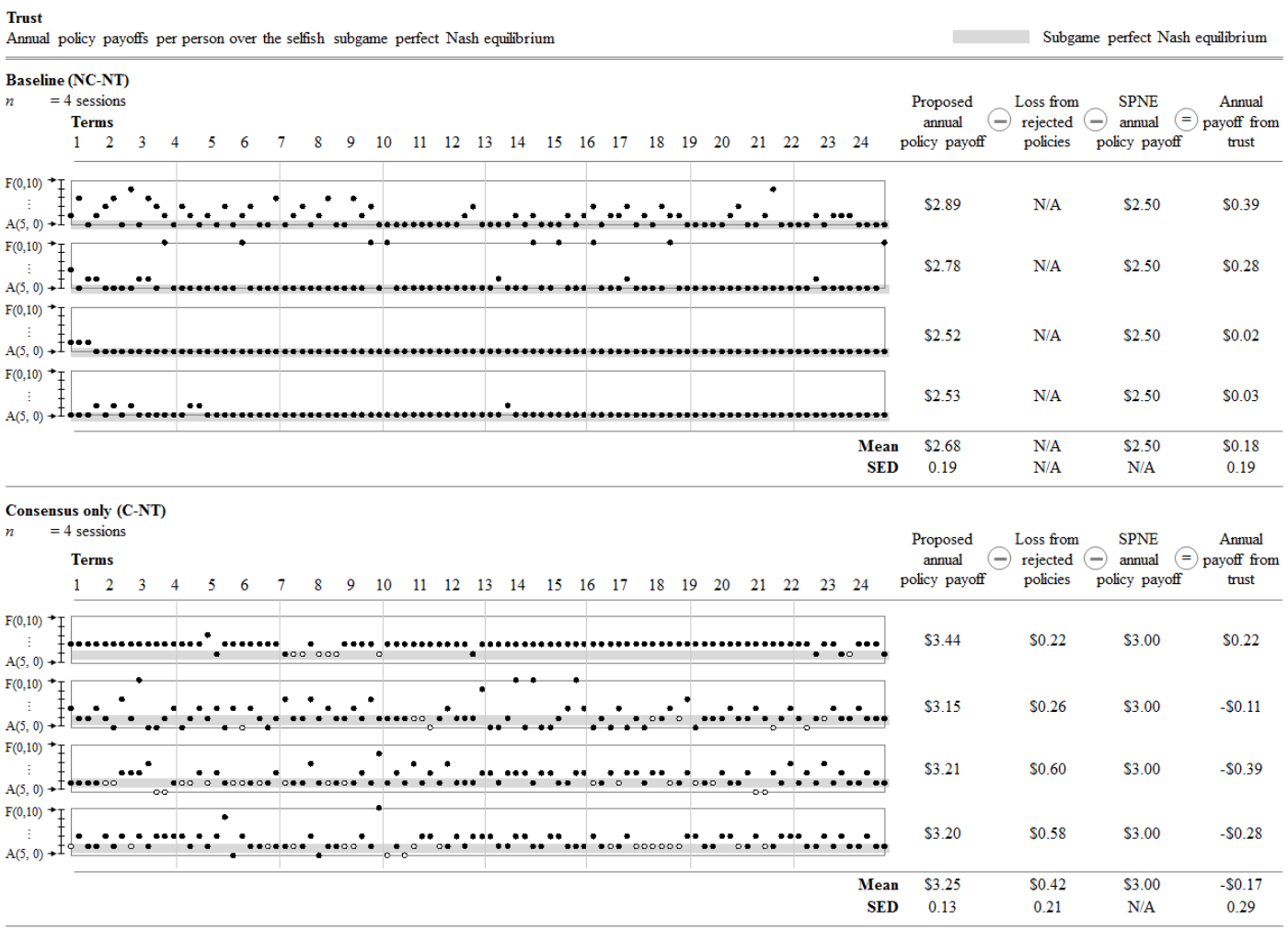
Budget for group 2	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	0.50	0.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	0.25	0.50	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	0.00	0.17	0.50	0.75	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	0.00	0.00	0.25	0.50	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	0.00	0.00	0.04	0.13	0.50	0.71	0.91	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	0.00	0.00	0.00	0.00	0.29	0.50	0.81	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	0.00	0.00	0.00	0.00	0.09	0.19	0.50	0.74	0.88	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00
7	0.00	0.00	0.00	0.00	0.02	0.04	0.26	0.50	0.76	0.91	0.97	0.99	1.00	1.00	1.00	1.00	1.00
8	0.00	0.00	0.00	0.00	0.01	0.01	0.12	0.24	0.50	0.69	0.87	0.93	0.97	0.99	1.00	1.00	1.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.09	0.31	0.50	0.78	0.88	0.94	0.99	0.99	1.00	1.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.13	0.22	0.50	0.62	0.78	0.89	0.94	0.97	0.98
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.07	0.12	0.38	0.50	0.71	0.87	0.92	0.97	0.98
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.06	0.22	0.29	0.50	0.69	0.81	0.88	0.93
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.11	0.13	0.31	0.50	0.67	0.78	0.88
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.06	0.08	0.19	0.33	0.50	0.62	0.77
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.12	0.22	0.38	0.50	0.71
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.07	0.12	0.23	0.29	0.50

Budget for group 1

Figure A2: Deriving Optimal Budget Distributions

Appendix 3

Figure A3: Experimental results, focusing on trust



Trust

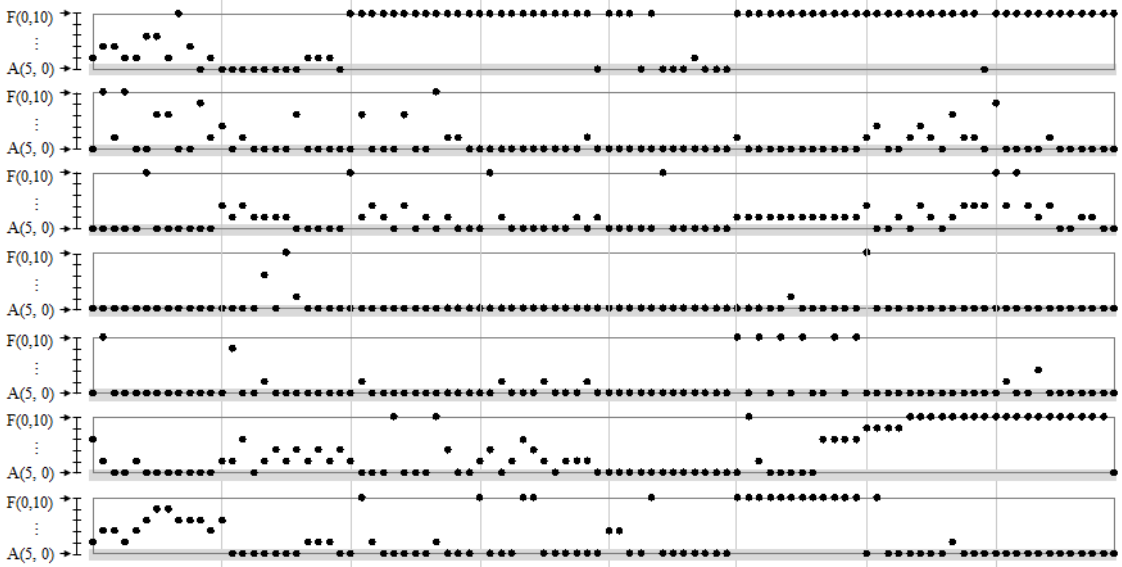
Annual policy payoffs per person over the selfish subgame perfect Nash equilibrium

Subgame perfect Nash equilibrium

Turn-taking only (NC-T)

n = 7 sessions

Terms



	Proposed annual policy payoff	− Loss from rejected policies	− SPNE annual policy payoff	= Annual payoff from trust
	$2.74	N/A	$2.50	$0.24
	$2.81	N/A	$2.50	$0.31
	$2.94	N/A	$2.50	$0.44
	$2.58	N/A	$2.50	$0.08
	$2.74	N/A	$2.50	$0.24
	$3.42	N/A	$2.50	$0.92
	$3.19	N/A	$2.50	$0.69
Mean	$3.13	N/A	$2.50	$0.63
SED	0.57	N/A	N/A	0.57

Figure A4: Experimental results, focusing on trust

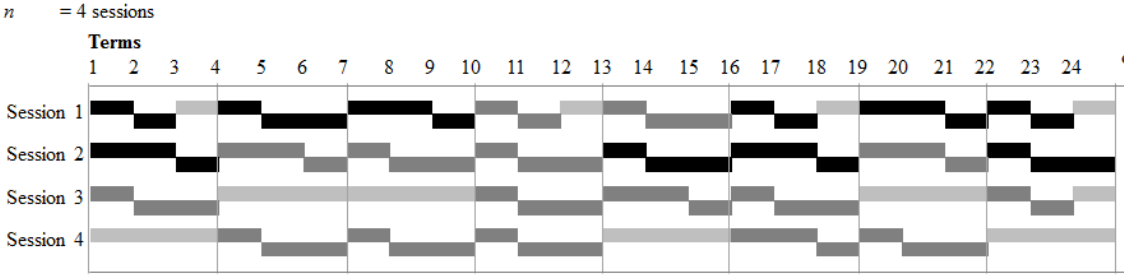
Figure A5: Experimental results, focusing on trust

Figure A6: Experimental results, focusing on adaptability

Adaptability
Number of realized trust opportunities

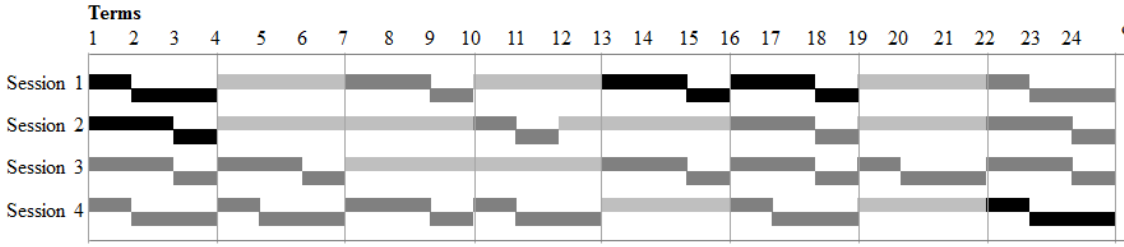
Trustor in office | *Trustee in office* | *Trust given* | *Trust returned* | *Either not given* | *Or not returned* | *No trustee*

Baseline (NC-NT)
n = 4 sessions



	Potential trust opportunities	(X)	Percent with trustee	(X)	Percent with trustee and mutual trust	(=)	Realized trust opportunities
Session 1	12		.67		.75		6
Session 2	8		1.0		.50		4
Session 3	9		.56		.00		0
Session 4	8		.63		.00		0
Mean	9.25		.71		.31		2.5
SED	1.89		.20		.38		3.0

Consensus only (C-NT)
n = 4 sessions



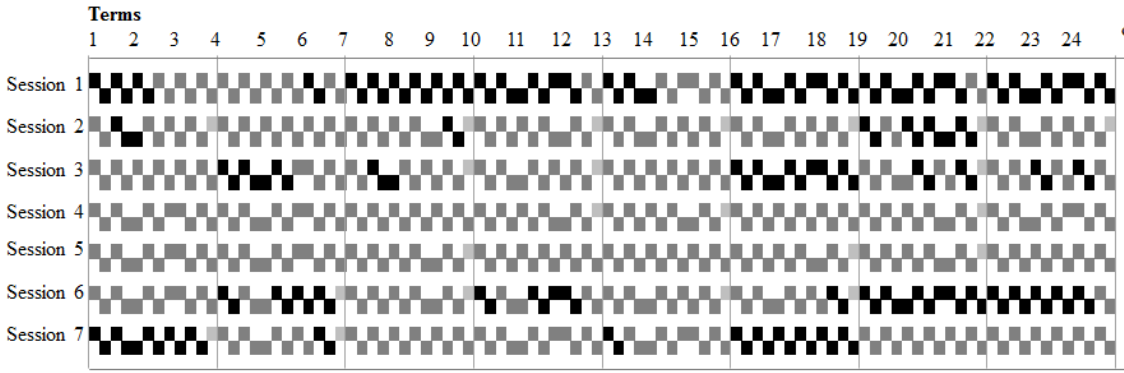
	Potential trust opportunities	(X)	Percent with trustee	(X)	Percent with trustee and mutual trust	(=)	Realized trust opportunities
Session 1	8		.63		.60		3
Session 2	9		.44		.25		1
Session 3	8		.75		.00		0
Session 4	8		.75		.17		1
Mean	8.25		.64		.25		1.25
SED	0.50		.14		.25		1.26

Figure A7: Experimental results, focusing on adaptability

Adaptability
Number of realized trust opportunities

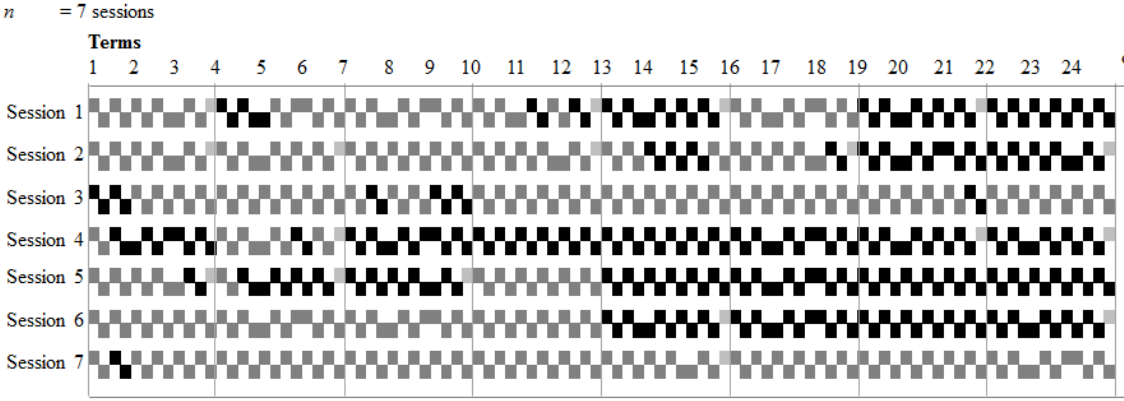
Trustor in office | *Trustee in office* | *Trust given* | *Trust returned* | *Either not given* | *Or not returned* | *No trustee*

Turn-taking only (NC-T)
n = 7 sessions



	Potential trust opportunities	(X) Percent with trustee	(X) Percent with trustee and mutual trust	(=) Realized trust opportunities
Session 1	43	1.0	.70	30
Session 2	48	.85	.15	6
Session 3	46	.93	.30	13
Session 4	45	.93	.00	0
Session 5	46	.93	.00	0
Session 6	45	.91	.44	18
Session 7	46	.95	.30	13
Mean	45.5	.93	.29	12.25
SED	2.08	.06	.30	12.97

Consensus and turn-taking (C-T)
n = 7 sessions



	Potential trust opportunities	(X) Percent with trustee	(X) Percent with trustee and mutual trust	(=) Realized trust opportunities
Session 1	45	.91	.49	20
Session 2	47	.89	.33	14
Session 3	48	1.0	.13	6
Session 4	45	.93	.88	37
Session 5	47	.94	.75	33
Session 6	45	.96	.49	21
Session 7	47	.98	.02	1
Mean	46.3	.93	.46	19.25
SED	1.5	.05	.32	13.15