Behavioral Spillovers and Cognitive Load in Multiple Games: An Experimental Study*

Jenna Bednar† Yan Chen‡ Tracy Xiao Liu§ Scott Page¶

November 1, 2010

Abstract

We present evidence from laboratory experiments of behavioral spillovers and cognitive load that spread across strategic contexts. In the experiments, subjects play two distinct games simultaneously with different opponents. We find that the strategies chosen and the efficiency of outcomes in one game depends on the other game that the subject plays in predictable directions. Using *entropy* as an empirical measure of behavioral variation in a normal form game, we find that games with low entropy have a stronger influence on behavior in the games with high entropy, and are less subject to influence by other games. Taken together, these findings suggest that people may not treat strategic situations in isolation but instead develop heuristics that they apply across games. Our findings suggest that behavior within a particular institution may depend upon other the incentive structures in play, and, as a result, institutional outcomes may be context dependent.

Keywords: multiple games, behavioral spillover, cognitive load, experiment
JEL Classification: C72, C91, D03

*This research was sponsored by grants from the National Science Foundation, the James S. McDonnell Foundation, and the United States Air Force Office of Scientific Research MURI. We are indebted to those who have supplied thoughtful reactions to our project, including Jasmina Arifovic, David Cooper, Vincent Crawford, Rachel Croson, John Duffy, Catherine Eckel, Simon Gächter, Ernan Haruvy, John Ledyard, Tom Palfrey, Rahul Sami, Rajiv Sethi, Roberto Weber, Kan Takeuchi, an anonymous referee, and seminar participants at the annual meeting of the Midwest Political Science Association (Chicago, March 2007), ICER (Turin, June 2007), the Third Asia Pacific Regional Meetings of the ESA (Shanghai, July 2007), the Santa Fe Institute (July 2007), USC Gould School of Law (November 2007), Shanghai Jiaotong University (June 2008), UT-Dallas (March 2008), the 2008 DFG-NSF Research Conference (New York, August 2008), the North America Regional Meetings of the ESA (Tucson, AZ, November 2008), the Allied Social Science Association Meetings (San Francisco, CA, January 2009). We would like to thank Andrea Jones-Rooy for excellent assistance.

†Department of Political Science, University of Michigan, Ann Arbor, MI 48109. Email: jbednar@umich.edu.

‡School of Information, University of Michigan, Ann Arbor, MI 48109. Email: yanchen@umich.edu.

§School of Information, University of Michigan, Ann Arbor, MI 48109. Email: liuxiao@umich.edu.

¶Center for the Study of Complex Systems, Departments of Political Science and Economics, University of Michigan, Ann Arbor, MI 48109. Email: scottepage@gmail.com.

1 Introduction

In this paper, we describe laboratory experiments in which individuals simultaneously and repeatedly play two games with different opponents. We test whether an individual's play in one game is influenced by the other strategic interaction she faces. Multiple games can increase cognitive load preventing individuals from choosing efficient or even equilibrium behaviors. They can also induce behavioral spillovers in which individuals choose similar strategies in the two games. We find evidence of both psychological processes.

First, we find that although individuals are free to apply distinct strategies in each game, they instead develop and apply common behaviors across the two games. For example, when the prisoner's dilemma is paired with a game of chicken, players are more likely to alternate on the off-diagonals rather than cooperate, compared to when the prisoner's dilemma is played alone or paired with other games. However, to our knowledge, there has not been a measure of game complexity which predicts when behavioral spillovers are present nor does any theory explain which games will influence others.

We introduce *entropy* of outcomes as a measure of behavioral variation in a normal form game, and posit that it is an empirical measure of the cognitive load induced by game complexity. For example, one of our games has a dominant strategy, efficient equilibrium. The play of this game produces the same outcome in almost every period, and therefore has low entropy. By our measure, this game is less cognitively taxing than the others. Using entropy of outcomes of play to organize our results, we find that games with low entropy have a stronger influence on behavior in the games with high entropy, and are less subject to influence by other games.[1]

To test for spillovers we posit specific hypotheses that are distinct from those that would be created by cognitive load alone. Cognitive load would imply that behavior varies significantly from control sessions when multiple games are played simultaneously but that the form of that variation is independent of the other game in the ensemble (provided that game demands equal cognitive attention). Yet, we find that behavior in one game depends significantly and predictably upon what other game is included in the two-game ensemble. This trend suggests that variance in actions cannot be attributed exclusively to cognitive load but instead indicates the presence of

[1] Our results are distinct from but complement earlier research on sequential behavioral spillovers, which have been interpreted as a form of priming or framing (Tversky and Kahneman 1986).

behavioral spillovers.

Our findings have an important implication for the study of games and for social science research more generally. If behavior in one game depends on other games an individual plays, then social scientists, whether doing experimental, theoretical, or empirical research, may need to consider the full ensemble of games that an individual faces (Samuelson 2001, Bednar and Page 2007). To date, almost all game theory research focuses on individual games, as do most experiments.

That norm is changing. A recent flurry of multiple game experiments demonstrates the existence and magnitude of ensemble effects. Our theory of behavioral spillovers can explain some of the findings in these experiments. Falk, Fischbacher and Gaechter (2010) investigate social interaction effects when two identical coordination or public goods games are played simultaneously with different opponents, and find no behavioral spillovers between the two games, which is consistent with our prediction that two identical games with the same entropy should not influence each other. Savikhin and Sheremeta (2010) study simultaneous play of a public goods game (low entropy) and a competitive lottery contest (high entropy). They find that cooperation in the public goods game reduces overbidding in contests, while contributions to the public good are not affected by the simultaneous participation in the contest. This result is consistent with our prediction that game influence flows from low entropy games to high entropy games. In a third study, using both a sequential and a simultaneous treatment, Cason, Savikhin and Sheremeta (2010) report cooperation spillovers from the median-effort game (low entropy) to the subsequent minimum-effort game (high entropy) when the games are played sequentially, but not simultaneously. Again, this result is consistent with our predictions based on entropy. Finally, Cason and Gangadharan (2010) investigate behavioral spillovers between a threshold public goods game and a competitive double auction market. They find that cooperation in public goods provision is less common when players simultaneously compete in the market.[2]

There do exist some sequential game experiments that find significant framing (Tversky and

[2]Huck, Jehiel and Rutter (2010) study feedback spillovers in sequential multiple games and find empirical support for an analogy-based expectation equilibrium (Jehiel 2005). They use a different protocol from the multiple game experiments discussed above. In their experiment, a player plays one of two games in each round, and sometimes receives the aggregate distribution of the play of the opponents over the two games, a design feature aimed to compare the long-run behavior in the presence and absence of feedback spillovers. Grimm and Mengel (2010) use a similar protocol where a player plays one of several games each round, and find that their data can be rationalized by either action- or belief-bundling.

Kahneman 1986) and learning transfer effects and are therefore supportive of our claim of behavioral spillovers (Cooper and Kagel 2008, Haruvy and Stahl 2010). For example, experiments that first auction off the right to play in the game produce different outcomes in subsequent games. In those experiments, the auction can be seen as an initial game. These pre-play auctions often lead to better outcomes. Van Huyck, Battalio and Beil (1993) demonstrate that without a pre-play auction, the median-effort coordination game played in isolation leads to inefficient equilibrium but that auctioning off the right to play before the coordination game leads to the payoff-dominant equilibrium. Crawford and Broseta (1998) explore the efficiency-enhancing effect of auctions theoretically using a model of stochastic, history-dependent learning dynamics, giving an analytic explanation for these results.[3]

The interest in multiple game experiments can be attributed to the ability to generate deeper insights into both individual and collective behavior. For learning about individual behavior, these experiments provide a laboratory in which subjects find themselves in a more cognitively taxing environment, one that resembles real world situations in which multiple stimuli simultaneously demand a person's attention.

At the collective level, the findings from multiple game experiments may contribute to an institutional explanation for behavioral variations in the play of common games. Distinct sets of experiences or cases lead distinct communities to draw different analogies when constructing strategies (Gilboa and Schmeidler 1995). Institutional interventions that take into account the behavioral repertoire of the relevant individuals may be more likely to succeed. Understanding behavior may be crucial to predicting institutional performance and designing institutions in different settings.

We have organized the paper as follows. In Section 2, we summarize the relevant theoretical literature and describe the specific games included in this study. Section 3 describes our experimental design. Somewhat unusually, in Section 4 we first present the results from the control sessions, where agents play a single game, and then develop our multiple-game hypotheses in Section 5, which are based on theory as well as results from the control sessions. Section 6 reports our findings on the ensemble effects. In Section 7, we discuss what these findings might mean and

[3]A additional strand of research that complements our findings looks at the emergence of cooperation. Weber (2006) reports the results of a minimum-effort game experiment where successful coordination is achieved in large groups by starting with small groups and adding entrants who are aware of the group's history. Successful coordination in large groups can be interpreted as learning transfer from small groups that find it easier to coordinate.

comment on potential future directions.

2 Theoretical Literature and the Games

In this section, we review the theoretical literature on multiple games and describe the specific games included in our study. Samuelson (2001) formally models behavioral spillovers and cognitive load when people play multiple games. He assumes that people pay a cognitive cost to analyze a strategic interaction. More sophisticated analyses require more cognitive load. In his model, individuals maintain a stock of analogies to organize their reasoning. In the analysis of three different bargaining games, the ultimatum game, the Rubinstein (1982) alternating offer bargaining game, and a tournament, he characterizes two equilibria, one in which the two bargaining games are played separately, and one in which they are played jointly. In the latter, players apply common analogies to disparate bargaining situations.

Samuelson's analysis is restricted to bargaining games. Bednar and Page (2007) examine behavioral spillovers and cognitive load effects in a broader class of six 2×2 games. They prove conditions for the existence and efficiency of behavioral externalities, using computational agent based models (Miller and Page 2007). Their agent based models show that simple learning rules could locate the proposed equilibria when played in isolation. When agents needed to solve multiple games simultaneously, the agents often created routines that they applied across strategic domains. The agent based model generates behavioral spillovers; agents employed identical strategies in distinct games.[4] The model also shows evidence of cognitive load: some ensembles of games outstrip the capacity of the agents to play each game optimally. In those cases, they find especially strong ensemble effects.

Given their focus on ensemble effects, Bednar and Page (2007) provide the main theoretical foundation for the current paper. Here we test whether the phenomena derived within models and generated by artificial agents can be produced in a laboratory with real people. We focus on four 2×2 games: the Prisoner's Dilemma (PD), Strong Alternation (SA), Weak Alternation (WA), and

[4]In comparison to the action-bundling results of Samuelson (2001) and Bednar and Page (2007), Jehiel (2005) uses a belief-bundling approach, where a player forms expectations about the behavior of the other players by pooling together several contingencies (analogy class) in which these other players must move, and forms an expectation about the average behavior in each analogy class. In his analogy-based expectation equilibrium, a player with coarser beliefs could still adopt different actions in different normal form games.

a Self Interest game (SI). The individual games belong to a class of two-person two-action games that contain a self-regarding action S and an alternative, C; in three of the games (PD, SA and WA), this alternative is cooperative. In these three games, cooperation lowers a player's own payoff and raises the payoff of the other and being selfish does the opposite, so in the one shot game, the unique dominant strategy equilibrium involves both players choosing selfish. In the fourth game, Self-Interest (SI), S is both the stage game dominant strategy and Pareto dominant.

The first game is a standard Prisoner's Dilemma, where the stage game has a dominant strategy equilibrium, (S, S), which is Pareto dominated by (C, C). Note that (C, C) also maximizes the joint payoff of the two players.

Prisoner's Dilemma: (PD)

	C	S
C	7, 7	2,10
S	10,2	4,4

In the second and third games, Strong Alternation (SA) and Weak Alternation (WA), while (S, S) remains the dominant strategy equilibrium for the stage game, agents do best (i.e., maximize joint payoff) in repeated play by alternating between the off-diagonals, (C, S) and (S, C). In Strong Alternation, the incentives to alternate are much stronger than in Weak Alternation. The alternation games are a distant cousin to the conventional Battle of Sexes and Game of Chicken, where agents are rewarded for coordinating their behavior. In our alternation games, four outcomes are rewarded with positive payoffs: CC, SS, and the alternating strategies of CS then SC and SC then CS. Coordinating on CC or SS is much less taxing than working out an alternating behavior, and the positive payoffs for each reduce the focality of an alternating equilibrium.

Strong Alternation: (SA)

	C	S
C	7, 7	4,14
S	14,4	5,5

Weak Alternation: (WA)

	C	S
C	7, 7	4,11
S	11,4	5,5

In the final game, Self Interest (SI), the dominant strategy equilibrium, (S, S), also Pareto dominates all other outcomes. Furthermore, in the stage game, S uniformly dominates C.

Self Interest: (SI)

	C	S
C	7, 7	2,9
S	9,2	10,10

These four games are variants of the six studied by Bednar and Page (2007). Our experiments with human subjects parallel the results of their computational experiments.

3 Experimental Design

Our experiments consist of four control sessions, each of which consists of a single game, and 14 treatment sessions, each of which consists of a pair of games. This experimental design enables us to determine the effects of ensemble on behavior by comparing the ensemble with the corresponding control sessions and to compare behavior across ensembles.

The control sessions follow the protocol of infinitely repeated games in the laboratory. We have one 12-player session for each of the single games. Participants are randomly matched into pairs at the beginning of each session, and play the same match for the entire experiment. In each session, participants first play the game for 200 rounds. After round 200, whether the game will continue to the next round depends on the "throw of the die" that is determined by the computer's random number generator. At the end of each round after round 200, with 90% chance, the game will continue to the next round. With 10% chance, the game stops. In other words, we implement an infinitely repeated game, with a discount factor of 1 for the first 200 rounds, and 0.9 thereafter. With the chosen discount factors, (C, C) can be sustained as a repeated game equilibrium in PD, SA and WA. With 12 players in each control session, we have 6 independent observations for each single game.

In the ensemble treatment, we again use twelve players in each session. Within each session, at the beginning, each player is randomly matched with two other participants, both of whom will be her matches for the entire experiment. She plays two distinct games with each of these people. This design allows us to analyze whether or not behavior in one game is influenced by the nature of the other game. As in the control sessions, we implement an infinitely repeated game, with a discount factor of 1 for the first 200 rounds, and 0.9 thereafter. Within each session, the 12 players are partitioned into independent groups of 4 each,[5] yielding 3 independent observations. As the two games are displayed side by side, we conduct two independent sessions for each game ensemble, changing the order of the display to avoid the order effect within each round. For example, for the

[5]The matching protocol is the following: $\underbrace{4-2-1-3}$, $\underbrace{6-5-7-8}$, $\underbrace{10-9-11-12}$ form three independent groups, each with four participants positioned on a circle, and each participant plays her left and right match.

game ensemble of SA and WA, we display SA as the left game in one session, and WA as the left game in another session. This way, if a player always makes decisions from left to right, we have a balanced number of observations for each order.

We used z-Tree (Fischbacher 2007) to program our experiments. As z-Tree does not record the mouse movements within each stage, we ran two additional sessions with ensembles, (SI, WA) and (WA, SI), where we use the software Morae to record the mouse movement. These two sessions enable us to determine the order of decisions within each round. The (SI, WA) session has 12 subjects, while the (WA, SI) has only eight subjects.[6]

Table 1: Features of Experimental Sessions

Control		Ensemble Treatment		
Game n	Pairs	(Left, Right)	n	Groups
		(PD, WA)	12	3
		(WA, PD)	12	3
		(PD, SI)	12	3
PD 12	6	(SI, PD)	12	3
		(SA, WA)	12	3
		(WA, SA)	12	3
		(SA, PD)	12	3
SA 12	6	(PD, SA)	12	3
		(SI, WA)	$12+12$	6
		(WA, SI)	$12+8$	5
		(SI, PD)	12	3
SI 12	6	(PD, SI)	12	3
WA 12	6			
Total 48	24		164	41

Table 1 reports features of experimental sessions, including the name of the game, the number of players in each session, the number of independent pairs for each control session, the ensemble of games, the number of players in each session, as well as the number of independent groups in each ensemble session.

Overall, 18 independent computerized sessions were conducted in the RCGD lab at the University of Michigan from March to October 2007, yielding a total of 212 subjects. Our subjects were students from the University of Michigan, recruited by email from a subject pool for eco-

[6]We recruited for twelve subjects, however, only eight showed up.

nomic experiments.[7] Participants were allowed to participate in only one session. Each ensemble treatment session lasted approximately 90 minutes, whereas each control session lasted about 45 minutes. The exchange rate was set to 100 tokens for $1. In addition, each participant was paid a $5 show-up fee. Average earnings per participant were $37.49 for those in the treatment sessions and $22.77 for those in the control sessions. Data are available from the authors upon request.

4 Results: Control Sessions

In this section, we report the results from the control sessions at the outcome level. This analysis provides a benchmark from which we can identify the presence of cognitive load and behavioral spillovers results in Section 6. In Subsection 6.2, we infer the repeated game strategies emerged in each game in the control and compare them with those in the ensembles. In this section, we treat each pair as an independent observation.

We first introduce an empirical measure of cognitive load. To measure the behavioral variation in a game, we apply the standard entropy concept to the outcome distributions.[8] The *entropy* of a random variable X with a probability density function, $p(x) = Pr\{X = x\}$, is defined by

$$H(X) = -\sum_{x} p(x) \log_2 p(x),$$

which is used to measure the amount of stochastic variation in a random variable that can assume a finite set of values. Therefore it is also a measure of the amount of information required to describe that distribution. When using logarithms to base 2, that measure is expressed in binary variables (bits).

For the analysis of two-person games, we model individual stage game strategies as a discrete random variable, X, with realizations in one of the four cells. Throughout the analysis, we use the convention that $0 \log 0 = 0$.[9] The entropy in a generic 2×2 game is in the interval $[0, 2]$, with the lower bound indicating certainty, i.e., all outcomes are in one cell, and the upper bound indicating a uniform distribution among the four cells. The cause of behavioral variation could be strategic uncertainty over what the other player will do.

[7]Graduate students from the Economics Department are excluded from the list.

[8]Shannon (1948) is credited with the development of the concept of entropy and the birth of information theory. Many basic concepts and findings in this field are summarized in Cover and Thomas (2006).

[9]This convention is easily justified by continuity, since $x \log x \to 0$ as $x \to 0$.

In Figures 1-4, we present time series data for each pair in each of the control sessions, with the entropy for each pair presented at the bottom of each graph.

[Figure 1 about here.]

Figure 1 presents outcomes in the Self Interest game. In this game, all six pairs converge to the Pareto dominant stage game equilibrium quickly and stay there. The entropy for each pair ranges from 0 to 0.04, indicating very little behavioral variation. This behavioral consistency is likely attributable to the uniform dominance property of the dominant strategy equilibrium in the stage game. Additionally, participants take an average of 0.62 seconds per round to make a decision in SI, significantly shorter than in any other game ($p \leq 0.01$, one-sided permutation tests). Based upon the uniform dominance property of the unique Pareto efficient stage game equilibrium, its low entropy, and the length of time it took for participants to complete the game, we posit that SI would be the easiest to play efficiently.

[Figure 2 about here.]

Figure 2 presents behavior in the Prisoner's Dilemma game. In this game, over half of the pairs play CC, the efficient outcome, which is consistent with findings from previous experiments (Andreoni and Miller 2002). Curiously, one pair also alternate for a fair number of rounds. The entropy for each pair ranges from 0.08 to 1.90, indicating changing behavioral variation. In addition, participants take an average of 1.00 second per round to make a decision in PD, significantly longer than SI, but shorter than SA ($p \leq 0.01$, one-sided permutation tests). As a "context" this game does not establish as strong a behavioral norm as the Self Interest game. Based upon this finding, we anticipate that PD will have a weaker behavioral pull than SI. The difficulty of learning to cooperate in the PD game may limit its spillover effects on play in other games.

[Figure 3 about here.]

Figure 3 presents behavior in the Strong Alternation game, where 5/6 of the pairs successfully establish the alternation outcomes. Pair 2 also attempts alternation on and off during the experiment. The entropy for each pair ranges from 1.29 to 1.84, indicating substantial behavioral variation.[10] In addition, participants take an average of 2.72 seconds per round to make a decision in SA, significantly longer than in any other game ($p \leq 0.01$, one-sided permutation tests). We interpret the longer decision time in SA as evidence that coordinated alternation requires more mental activities to establish. Since successful alternation is established in five out of six pairs, this game also provides a strong context.

[Figure 4 about here.]

Last, Figure 4 presents the dynamics from the Weak Alternation game. In this game, only two out of six pairs develop an alternating behavior, two pairs cooperate, one (pair 4) converges to SS, and the last pair (pair 6) does not seem to have converged to a stable outcome. The entropy for each pair ranges from 0.44 to 1.91, with the highest aggregate entropy among all four games. In addition, participants take an average of 1.24 seconds per round to make a decision in WA, significantly longer than SI, shorter than SA ($p \leq 0.01$, one-sided permutation tests), not significantly different from PD ($p = 0.138$, one-sided permutation test). As WA results in higher behavioral variation, we speculate that, while subject behavior in WA is more likely to be influenced by the other game in an ensemble, when paired with other games, it might increase the subjects' cognitive load.

Table 2: Distribution of Outcomes and Entropy in Control Sessions

	SI		PD		SA		WA	
	C	S	C	S	C	S	C	S
C	0.00	0.14	55.68	11.67	5.02	39.81	33.18	21.57
S	0.00	99.86	14.82	17.82	40.37	14.81	22.74	22.51
Entropy	0.02		1.68		1.68		1.98	

To summarize our findings, Table 2 reports the aggregate distribution of outcomes in each of the four games in the control sessions, and the respective entropy for each game in the last line. The behavioral variation measured by entropy is the lowest in the Self Interest game (0.02), followed by Prisoner's Dilemma and Strong Alternation (1.68), and the highest in Weak Alternation (1.98). We postulate that games with lower behavioral variation (entropy) might have stronger behavioral spillovers than those with higher behavioral variation.

[10] Perfect coordinated alternation results in an entropy of 1.

In our outcome level analysis, we focus on the Pareto efficient outcomes: (1) for both players to always play selfish each round (SS) in the Self Interest game; (2) for both players to always cooperate (CC) in Prisoner's Dilemma; and (3) coordinated alternation between S and C (ALT) in the Strong and Weak Alternation games, respectively. These outcomes also coincide with the simplest equilibria among the set of Pareto efficient ones in our set of games.[11] Table 3 reports the proportion of Pareto efficient outcomes in each game over the entire series, and the corresponding p-values for the one-sided permutation tests for each pairwise comparison. Boldfaced numbers are the mode of the distribution.

Table 3: Average Proportion of Pareto Efficient Outcomes in Control Sessions

	% Pareto Efficient Outcomes			P-value of Permutation Tests		
Games	SS	CC	ALT	CC v. SS	CC v. ALT	SS v. ALT
SI	**99.86**	0.00	0.00	0.000	0.500	0.000
PD	17.82	**55.68**	15.44	0.039	0.031	0.389
SA	14.81	5.02	**71.12**	0.040	0.000	0.001
WA	22.51	33.18	**36.14**	0.308	0.430	0.317

In the control sessions, the proportion of Pareto efficient outcomes (SS in SI, CC in PD, ALT in SA and WA) is significantly higher than any other outcomes in SI, PD and SA ($p < 0.05$). However, in WA, there is no significant difference in the proportions of any of the three outcomes ($p > 0.10$).

In sum, three distinct outcomes emerge in the control sessions: in the SI game, selfishness; in the PD game, cooperation; and in SA and WA an alternating form of cooperation, where subjects alternate between the cooperative and selfish actions. Weak Alternation has weaker incentives, so the coordinated alternation is not as prominent as with Strong Alternation.

5 Hypotheses of Ensemble Effects

In this section, we present a set of hypotheses testing the null of game independence against our two posited ensemble effects, behavioral spillovers and cognitive load. Our alternative hypotheses

[11] When we represent a repeated game strategy as an automaton, the simplest strategy is defined as one with the least number of states (Kalai and Stanford 1988, Baron and Kalai 1993). We present our repeated game strategy analysis in Subsection 6.2.

are based on the theoretical results from Bednar and Page (2007), as well as our empirical results from the control sessions presented in Section 4. These hypotheses are also broadly consistent with the analogy based model of Jehiel (2005) and the case based reasoning of Gilboa and Schmeidler (1995).

The general null hypothesis in our investigations is of game independence: play in one game is not be affected by the existence of another game to play. If the independence hypothesis is correct then we should see no difference between behaviors in the control sessions (games played in isolation) and when games are presented to subjects as part of ensembles.

Our experimental design tests the existence of two types of ensemble effects: *behavioral spillovers* and *cognitive load*. If games presented within ensembles create *behavioral spillovers*, subjects will respond as if they are developing heuristics that they apply across games. In particular, dominant behavior in one game will influence choice in another. With cognitive load effects, subjects might resort to the dominant strategy in the stage game more often when a game is paired with another game with higher behavioral variation, such as the Weak Alternation game.

These effects may be manifested when we compare behavior in a game when it is matched with different games. Our investigations center on whether the ensemble play depend upon *which* other game is in the ensemble. A positive answer to this question would fail to support the null hypothesis of independent play, and support instead the hypothesized contextual dependence of game play.

The following alternative hypotheses look at behavioral spillovers. The general alternative hypothesis is that subject's choice of action in a particular game will be influenced by the other game in the ensemble, particularly biasing choice toward the other game's simple Pareto optimal strategy. Specifically, we expect:

Hypothesis 1 (Effect of SI)**.** *Compared to other ensembles, games paired with Self-Interest will exhibit more selfishness.*

Hypothesis 2 (Effect of PD)**.** *Compared to other ensembles, games paired with the Prisoner's Dilemma will exhibit more cooperation.*

Hypothesis 3 (Effect of SA)**.** *Compared to other ensembles, games paired with Strong Alternation will exhibit more alternation.*

As the Weak Alternation game only generates 36% of alternation in the control sessions, we do not expect that games paired with Weak Alternation will exhibit more alternation. Instead, we expect that it might have an effect in terms of cognitive load.

Our hypotheses relating to cognitive load begin with an obvious one for which we find overwhelming support: we expect efficiency to fall in the multiple game settings unless the game is easy to play.

Hypothesis 4 (Multiple Games and Cognitive Load)**.** *Compared to the corresponding control sessions, subjects will produce less efficient behavior in any game that requires a non-trivial cognitive load.*

In Section 4, we develop a partial ordering of the four games based upon the entropy of each game in the control sessions, i.e., the behavioral variation follows the order of SI $<$ PD $\sim$ SA $<$ WA. Based on the entropy, decision time and the payoff structure of each game, we posit that Self Interest is the only easy game to play so it will be the only game for which we do not expect to see a significant falloff in efficiency.

According to our measure, Weak Alternation is the most difficult, and this relative difficulty will be reflected in subjects' choice of action. In particular, we conjecture that cognitive load effects will be most prevalent in games played with more difficult games.

Hypothesis 5 (Effect of WA: Cognitive Load)**.** *Compared to other ensembles, subjects will exhibit more selfish behavior (stage game dominant strategy) in a game when it is paired with WA.*

Hypothesis 5 conveys our interest in the experiment's ability to reveal limitations in the cognitive processing of subjects playing multiple games simultaneously. While there are no design features to the experiments that would preclude the subjects from optimizing in each game, we believe that when subjects are asked to solve two games simultaneously they will not be as efficient as they are when playing an isolated game. In particular, when an ensemble contains WA, we predict more *selfish* behavior in PD or SA, a stage-game dominant strategy but inefficient in PD or SA.

6 Results: Ensemble Effects

In this section, we present ensemble effects at the outcome level (subsection 6.1) as well as those at the strategy level (subsection 6.2). In all our analysis in this section, a pair in a control session or a group of four in an ensemble session is treated as an independent observation.

6.1 Ensemble Effects at the Outcome Level

Our anticipation was that subjects would play particular games differently between the control sessions, where they played a single game, and when that game appeared as part of an ensemble. This prediction emerges from the two core hypotheses: both behavioral spillovers and cognitive load will affect play in ensembles. Consequently, we expect different outcomes between the control sessions and the corresponding ensembles. In the control sessions, the Pareto-efficient outcomes (SS in SI, CC in PD, ALT in SA and WA) emerge as the mode among all the three outcomes in SI, PD and SA played in isolation. We first examine the likelihood of Pareto-efficient outcomes when a game is played in an ensemble.

Table 4: Pareto Efficient Outcomes in Ensemble vs. Control

	SS in SI	CC in PD	ALT in SA	ALT in WA	Pooled PD, SA, WA
% in Control	0.999	0.557	0.711	0.361	0.541
% in Ensemble	0.991	0.410	0.449	0.244	0.360
H_1: control > ensemble one-sided p-values	0.002	0.171	0.031	0.213	0.034
# of observations (control, ensemble)	(6, 23)	(6, 18)	(6, 18)	(6, 23)	(18, 59)

Note: In the permutation tests, we treat each pair (group of four) in the control (ensemble) sessions as one observation.

Table 4 reports the proportion of Pareto efficient outcomes in the control and ensembles, as well as p-values from one-sided permutation tests. The general null hypothesis is game independence, i.e., the proportion of Pareto efficient outcomes of a particular game is the same between the control and ensemble treatments, while the alternative hypothesis is that the the proportion of Pareto efficient outcomes is higher when a game is played alone. We summarize the results below.

Result 1 (Pareto Efficient Outcomes in Ensembles)**.** *Averaging across all games except SI, the proportion of Pareto efficient outcomes decreases by 18% when the games are part of ensembles. An individual game analysis shows a decrease in Pareto efficient outcomes for all games and a significant decrease for ALT in SA.*

Support. *In Table 4, we reject the null of game independence in favor of H_1 for SS in SI ($p = 0.002$), ALT in SA ($p = 0.031$), and the pooled PD, SA and WA games ($p = 0.034$).*

Result 1 indicates that, in three out of four games, behavior in a game played in an ensemble is different from the same game played in isolation. In particular, in Strong Alternation, subjects alternate significantly less ($p = 0.031$) when it is in an ensemble than when it is played in isolation. The Self Interest game, however, is not affected by the presence of other games. It is the easiest game to play. Whether in control or ensemble treatments, subjects quickly converge to SS, with over 99% of selfishness across the rounds. Thus, by Result 1, we reject the null of game independence in favor of Hypothesis 4 in SI, SA and the pooled PD, SA and WA games.

To establish the existence of behavioral spillovers in the presences of cognitive load, we compare outcomes between ensembles. We first examine how outcomes in the Weak Alternation game are influenced by other games in the ensemble. Recall that Weak Alternation has the greatest behavioral variation (entropy) in the control sessions. This feature should render it susceptible to influence from other games in the ensemble.

We focus first on play in Weak Alternation in which the other games in the ensemble are Strong Alternation and the Prisoners' Dilemma. These two games create similar behavioral variation in our control sessions (entropy $= 1.68$ in each game), so we posit that their effects on cognitive load will be similar. If behavioral spillovers are present in the Weak Alternation, then we should see more alternation in that game when it is paired with Strong Alternation, and we should see more cooperation when it is paired with the Prisoners' dilemma. Both effects are present.

Result 2 (Behavioral Spillover to WA)**.** *Comparing (**WA**, PD) and (**WA**, SA), subjects alternate more in WA when also playing SA (37% versus 18%) and cooperate significantly more when WA is paired with PD (31% versus 11%).*

Support. *One-sided permutation tests comparing the proportion of ALT in (**WA**, PD) and (**WA**,* *SA) yield* $p = 0.12$*. Similarly, one-sided permutation tests comparing the proportion of CC in WA between the two ensembles yield* $p = 0.022$*.*

Result 2 indicates that behavior in Weak Alternation is indeed susceptible to the influence of the other games in the ensemble: participants alternated more in WA when also playing SA and cooperated more when WA was paired with PD. By Result 2, we reject the null of game independence in favor of Hypotheses 2 and 3.

We next look for behavioral spillovers in the PD game. Here, we find more ALT when PD is paired with SA, which supports spillovers. But even more compellingly, we find that when PD is paired with SI, we get twice as much SS in the PD game, which is consistent with a hypothesis of behavioral spillovers. Given that SI is the least cognitively tasking of the games (entropy $= 0.02$), this finding cannot be explained by cognitive constraints alone.

Result 3 (Behavioral Spillover to PD)**.** *Comparing (**PD**, SA) and (**PD**, SI), we observe more ALT in PD when it is paired with SA (21% versus 5%), and more SS in PD when it is paired with SI (46% versus 23%). Comparing (**PD**, SA) and (**PD**, WA), we observe more ALT in PD when it is paired with SA (21% versus 9%) and significantly more SS in PD when it is paired with WA (39% versus 23%).*

Support. *One-sided permutation test comparing the proportion of ALT in PD in ensembles (**PD**, SA) and (**PD**, SI) yields* $p = 0.10$*, while one-sided permutation test comparing the proportion of SS in PD between the two ensembles yield* $p = 0.10$*. Similarly, one-sided permutation test comparing the proportion of ALT in PD in ensembles (**PD**, SA) and (**PD**, WA) yields* $p = 0.18$*, while one-sided permutation test comparing the proportion of SS in PD between the two ensembles yields* $p = 0.04$*.*

Result 3 indicates that the Prisoner's Dilemma is susceptible to the institutional context it is situated in. While participants alternate more in PD when they also play Strong Alternation, they play PD selfishly more when they also play Self Interest or Weak Alternation. Thus, we reject the null in favor of Hypotheses 1, 3 and 5.

Finally, we consider the effect on the SA game. Here, we compare play in SA when it is paired with SI and with PD. Clearly, SI is the least cognitively tasking of the games. Therefore, if cognitive load were the only force operating, we would expect higher efficiency in the SA game when it is paired with SI. However, we find that pairing SA with SI results in the same efficiency as when SA is paired with PD. This can be explained by the fact that pairing SA with PD produces more cooperation. This provides further evidence of behavioral spillovers.

Result 4 (Behavioral Spillover to SA)**.** *Comparing (**SA**, PD) and (**SA**, SI), we observe CC more often in SA when paired with PD (15% versus 7%). The difference is more pronounced and significant in the second 100 rounds (13% versus 1%).*

Support. *One-sided permutation tests comparing the proportion of CC in in SA in ensembles (**SA**, PD) and (**SA**, SI) yield* $p = 0.11$ *for the entire series, and* $p = 0.013$ *for the second 100 rounds.*

In later rounds, the effect grows even more pronounced: while they continue to cooperate in SA when it is paired with PD, where SA is paired with SI subjects shift from CC to SS, so that the CC percentages are 13% in (**SA**, PD) versus 1% in (**SA**, SI). By Result 4, we reject the null in favor of Hypothesis 2.

We next compare the efficiency generated in each game. Following convention, we capture efficiency by the percentage of potential payoff above the minimum payoff that the players receive. Our normalized efficiency measure is defined as follows.

$$\text{Efficiency} = \frac{\text{Actual joint payoffs} - \text{Minimum joint payoffs}}{\text{Maximum joint payoffs} - \text{Minimum joint payoffs}} \tag{1}$$

Table 5: Efficiency in the Control and Ensemble

Control	Efficiency in Control	Ensemble (Game1, Game2)	Average Efficiency in Ensemble		
			Game1	Game2	Ensemble
		(SI,PD)	99.32	51.50	80.19
		(SI,SA)	99.23	64.24	82.76
SI	99.86	(SI,WA)	98.99	63.06	86.16
PD	73.35	(PD,SA)	63.59	68.10	66.17
SA	82.69	(PD,WA)	53.97	53.97	53.97
WA	70.86	(SA,WA)	55.66	57.89	56.52

Table 5 presents the average efficiency in the control and the ensemble sessions. For each ensemble, we present the efficiency of each game in the ensemble, as well as the overall ensemble efficiency. Consistent with Result 1, the efficiency in SI, PD and SA control sessions is higher than that of the corresponding games in ensemble sessions. In particular, the following comparisons are significant at the 5% level: SI > (**SI**, SA), SI > (**SI**, WA), and SA > (**SA**, WA).

Table 6: P-values of Pairwise Efficiency Comparisons between Ensembles

Ensemble	(SI,SA)	(SI,WA)	(PD,SA)	(PD,WA)	(SA,WA)
(SI,PD)	0.366	0.114	0.079	0.004	
(SI,SA)		0.770	0.039		0.022
(SI,WA)				0.000	0.003
(PD,SA)				0.079	0.227
(PD,WA)					0.416

Table 6 reports the p-values of one-sided permutation tests of pairwise comparison between ensembles which have one game in common. We find that a game played together with SI generates significantly higher efficiency than the corresponding game paired with any other game. Specifically, the following comparisons are significant at the 5% level, except for (SI, PD) $\geq$ (SA, PD), which is significant at the 10% level:

- (SI, PD) > (WA, PD) and (SI, PD) $\geq$ (SA, PD);
- (SI, SA) > (PD, SA) and (SI, SA) > (WA, SA);
- (SI, WA) > (PD, WA) and (SI, WA) > (SA, WA).

The efficiency comparison across ensembles follows from several behavioral regularities. First, unlike any other game, the Self Interest game generates consistently high efficiency regardless of whether it is play alone or with other games. Second, SI takes significantly less time to play than any other game, indicating lower cognitive load, which implies that participants can devote more time to optimize in the other game which causes overall high efficiency in the ensemble.

In general, the experimental results agree with our alternative hypotheses: game independence is not supported, but instead subjects are influenced by contextual effects of behavioral spillovers and cognitive load. While analyses reported in this subsection are based on non-parametric tests, we recognize that income effects might influence behavior. Thus, we use probit regressions with standard errors clustered at the pair (respectively group) level for Results 1 to 4, controlling for income and learning effects in each specification. Following Ham, Kagel and Lehrer (2005), we use a participant's cash balance prior to period t to control for income effects. We find that, while cash balance is statistically significant in most specifications, its marginal effect on behavior is never greater than 1%.[12] Further, Results 1 to 4 continue to hold in our regression analysis.

6.2 Ensemble Effects at the Strategy Level

In this subsection, we analyze repeated game strategies in each game, and the ensemble effects at the strategy level. Following Rubinstein (1986) and Abreu and Rubinstein (1988), we use automaton (or the Moore machine) to represent repeated game strategies. In Figure 5, we present 29 repeated game strategies, including the exhaustive set of 26 1- and 2-state automata, and three 3-, 4- and 5-state automata which have been widely discussed in previous literature (Engle-Warnick and Slonim 2006). For each subject i, we calculate the *fitting proportion* for each automaton, M_j, for the entire sequence of observed actions $\{a_i^t\}_{t=1}^T$, defined as $F_i(M_j) = \sum_{t=1}^T I(a_i^t, M_j^t)/T$, where the indicator function $I(a_i^t, M_j^t) = 1$ if $a_i^t = M_j^t$. At the treatment level, we define the average fitting proportion as $\bar{F}(M_j) = \sum_{i=1}^n F_i(M_j)/n$, where n is the number of players.

To highlight the main results, in subsequent discussions, we only compare the performance of a subset of 1- and 2-state automata.[13] Table 7 presents 9 out of 26 1- and 2-state automata, including all six strategies with at least 50% fitting proportion across all games and ensembles, and three which do not survive the 50% fitting threshold but have been extensively discussed in the literature (Engle-Warnick and Slonim 2001; Bednar and Page 2007; Hanaki et al. 2005). These 9 strategies can be divided into three categories: (1) cooperative strategies, including Always Cooperate (AC), Forgive Once (F1), Suspicious Forgive Once (sF1); (2) reciprocal strategies, including Tit-for-Tat (TFT), and Suspicious Tit-for-Tat (sTFT); and (3) selfish strategies, including Always Selfish (AS), Grim Trigger (GT), Switch after Cooperate (SAC), suspicious Switch after Cooperate (sSAC). Of these nine strategies, AC, AS and GT do not survive the 50% fitting threshold.

The strategy level analysis in the control sessions reveals few surprises. In the PD game, reciprocal strategies, TFT and sTFT, have significantly higher fitting proportion, $\bar{F}(TFT) = \bar{F}(sTFT) = 0.84$, than every other strategy.[14] This result is consistent with the literature where

[12] Regression tables are available from the authors upon requests.

[13] The complete analysis of all 29 automata are available from the authors upon request.

[14] $\bar{F}(TFT) > \bar{F}(M_j)$: $p < 0.05$ for any $M_j \neq \{AC, sTFT\}$, and $p < 0.10$ for $M_j = AC$. Similarly, $\bar{F}(sTFT) > \bar{F}(M_j)$: $p < 0.05$ for any $M_j \neq \{sF1, AC, TFT\}$, and $p < 0.10$ for $M_j \in \{sF1, AC\}$, one-sided Wilcoxon signed rank tests.

Table 7: Description of Nine Strategies

Strategy Category	Name of Strategy	Strategy Number	Initial Action	Continued Play
Cooperative	Always Cooperate (AS)	M1	C	Always play cooperate
	Forgive once (F1)	M8	C	Goes to S if other plays S and goes to C when the last period is S
	Susipicious Forgive once(SF1)	M9	D	Goes to S if other plays S and goes to C when the last period is S
Reciprocal	Tit for Tat (TFT)	M4	C	Copy other’s previous action
	Susicious Tift For Tat (STFT)	M5	S	Copy other’s previous action
Selfish	Always Selfish (AS)	M2	D	Always play selfish
	Grim Trigger (GT)	M3	C	C until other play S, then S forever
	Switch After C(SAC)	M6	C	After C, play S until other plays C
	Susipicious Switch After C(SSAC)	M7	S	After C, play S until other plays C

TFT is widely and successfully used in the repeated PD simulations (Axelrod and Hamilton 1981, Bednar and Page 2007, Hanaki et al 2005). These two strategies have even better fitting proportion, $\bar{F}(TFT) = \bar{F}(sTFT) = 0.90$, in the Weak Alternation game.[15] In the Strong Alternation game, TFT, sTFT as well as SAC and sSAC are all best fitting strategies, each with a fitting proportion of $\bar{F} = 0.85$, and each significantly outperforming every other strategy outside this set.[16] This result is consistent with Hanaki et al (2005) where in a simulation of coordination games similar to Strong Alternation, SAC is the best performing strategy. Finally, in the Self Interest game, due to the lack of variation in actions, player behavior can be explained by any different strategy that produces all selfish behavior. Among these strategies are TFT, sTFT, SAC, sSAC, AS, and GT, each has a fitting proportion of $F = 1.00$.

The strategy level analysis for the ensemble sessions also aligns with expectations. Comparing strategies used in PD when it is paired with SI and those used in the PD control sessions, we find that the fitting proportion of selfish strategies increases from control to ensembles. For example, the fitting proportion of strategies that choose to be selfish after cooperating (SAC and sSAC) increases from 0.59 in the control to 0.71 in the ensemble ($p = 0.08$, one-sided permutation test). It is still the case though that TFT and sTFT retain the greatest fitting proportion in PD ($\bar{F} = 0.91$) even when it is paired with SI. Likewise, when PD is paired with WA, the fitting proportion of selfish strategies again increases, with $\bar{F}(AD) = 0.31$ in the control sessions and 0.49 in the (PD, WA) ensemble ($p = 0.09$), $\bar{F}(SAC) = 0.59$ in the control and 0.66 in the ensemble ($p = 0.06$), and $\bar{F}(sSAC) = 0.59$ in the control and 0.66 in the ensemble ($p = 0.07$).

Comparing strategies used in SA when it is paired with SI and those used in the SA control sessions, we find that the fitting proportion of selfish strategies significantly increases from the control to the ensemble, while that of cooperative strategies weakly decreases. Specifically, $\bar{F}(GT) = 0.55$ in the control and 0.65 in the ensemble ($p = 0.02$), while $\bar{F}(AC) = 0.45$ in the control and 0.37 in the ensemble ($p = 0.10$). We obtain similar results comparing the (SA, WA) ensemble and SA control sessions.

[15] $\bar{F}(TFT) > \bar{F}(M_j)$: $p < 0.05$ for any $M_j \neq \{SAC, sSAC, sTFT\}$, and $p < 0.10$ for $M_j \in \{SAC, sSAC\}$. Similarly, $\bar{F}(sTFT) > \bar{F}(M_j)$: $p < 0.05$ for any $M_j \neq \{F1, SAC, sSAC, TFT\}$, and $p < 0.10$ for $M_j \in \{F1, SAC, sSAC\}$, one-sided Wilcoxon signed rank tests.

[16] SAC or sSAC has significantly greater fitting proportion than any other strategy outside the set ($p < 0.05$), while TFT or sTFT has significantly greater fitting proportion than any other strategy outside the set ($p < 0.05$) except for F1 ($p < 0.10$), one-sided signed rank tests.

Last, when we compare repeated game strategies between ensembles, several of the results from the outcome level analysis survive, albeit in a slightly different form. For example, comparing behavior in WA in the (WA, PD) vs.(WA, SA) ensembles, we find that, when WA is paired with PD, the fitting proportion of cooperative strategies, such as always cooperate (AC), is weakly higher than when it is paired with SA (0.45 vs. 0.36, $p = 0.10$), while that of selfish strategies, such as AD (55 vs. 64%), GT (55 vs. 64%), SAC (72 vs. 81%) and sSAC (72 vs. 81%) increases ($p < 0.1$ for each comparison, one-sided permutation tests).

In sum, analysis of behavioral spillovers at both the outcome and strategy levels yields largely consistent results, i.e., when games are paired in ensembles, play differs from isolated controls, and in predictable ways. In some cases, the two levels of analysis provide different lenses on the same phenomenon. For example, both analyses demonstrate that the SI game leads to more selfish behavior and less cooperative behavior. In other cases, the strategy level analysis highlights a different feature of the results. For example, the strategy analysis allowed us to see how often pairs were able to coordinate on alternating strategies. A thorough reading of the data at both levels – outcomes and strategies – reveals unequivocal evidence of behavioral spillovers and cognitive load.

7 Discussion

In this paper, we present an experimental study to test for ensemble effects in game playing behavior. We test for these effects looking both at outcomes and strategies. Our study reveals strong evidence of behavioral spillovers that depend in predictable ways on features of the games in the ensemble. In particular, if subjects play one game in an ensemble that encourages selfishness or cooperation, then they are more likely to exhibit that behavior in the other game in their ensemble, even though they play the other game with a different player. We also see evidence of cognitive load.

To derive our hypotheses about cognitive load and behavioral spillovers, we introduce a measure of behavioral variance, *entropy*. We posit that in ensembles that include games that produce high entropy outcomes, cognitive load will be most pronounced. Consistent with our expectations, cognitive load has the greatest effect when ensembles include Weak Alternation, our highest-entropy game. In contrast, we hypothesize that low entropy games would produce stronger behavioral spillovers and are less influenced by other games. Both predictions are supported by our laboratory findings.

Our findings provide an initial demonstration of how a person's behavior in a given game depends on the ensemble of strategic situations that the person faces. In doing so, they call into question the focus on isolated games in most theoretical and empirical analyses. This critique extends to mechanism design, which assumes that incentives can be considered independent of the broader behavioral context.

Evidence of behavioral spillovers may shed light on policy choices. Decades of social science research show that policies geared toward the political and economic improvement of developing nations often fail. Interventions that appear efficient on paper play out quite differently on the ground (Easterly 2006). Near identical institutions implemented in multiple contexts often produce divergent outcomes. Putnam (1993) chronicles the disparate performance of an identical institutional innovation in northern and southern Italy. To explain this unpredictability of institutional interventions, many scholars have focused on belief systems and trust relationships and how they support or fail to support incentive structures (e.g. North (2005), Grief (2006), Putnam (1993)). The core argument of these papers is that institutions support multiple equilibria, some good and some bad, and that the good equilibria require particular beliefs.

Our findings suggest an alternative explanation, namely that people possess behavioral repertoires that they carry from one strategic situation to the next. Behavior in one context affects behavior in others (Gigerenzer and Selten (2002), Page (2007), Bednar and Page (2007)). It follows that how an institution fares in any particular setting should depend on those repertoires. The behavioral repertoire approach complements the belief based approach when thinking about policy. An institutional intervention may require high levels of trust to be successful. It may also require particular behavioral skills: the ability to bargain, or the ability to enact second order punishment. Behaviors in the community, in addition to beliefs about trustworthiness, may be relevant data informing evaluations of institutional performance. Behaviors take on even greater importance when designing institutions. If individuals apply existing behaviors to new contexts, then institutions can benefit by leveraging existing behaviors, thereby reducing behavioral uncertainty (Gilboa and Schmeidler 1995).

As with any laboratory experiment, our results may not translate to the larger world. Outside the laboratory, people rely on contextual clues to behave differently in distinct situations. Thus, people can act altruistically toward their children but competitively at work. We do not deny the human capacity to bracket contexts and act accordingly. However, we believe that such contextual bracketing requires cognitive effort and that, in general, people will seek out consistent behaviors that apply across multiple settings. Thus, bracketing may pull in the opposite direction, but the force would have to be substantial to overwhelm the drive toward consistency we see here. A second potential criticism pertains to the simplicity of the games we consider. Would these effects continue to hold for more complex games embedded in a richer institutional context? We cannot answer that question with this set of experiments. But the fact that the game ensembles can influence behaviors in individual games in a relatively sterile laboratory would seem to suggest that such effects might also exist in the real world. Furthermore, recent multiple game experiments using more complex games demonstrate cognitive load and behavioral spillover effects largely consistent with our findings (e.g., Savikhin and Sheremeta 2010, Cason, Savikhin and Sheremeta 2010).

To summarize, these experiments demonstrate that significant ensemble effects emerge in the laboratory setting. And, more importantly, these ensemble effects can be predicted based upon the the attributes of the games. Outcome level and strategy level analysis show consistent ensemble effects. Subjects with incentives to behave cooperatively (resp. selfishly) in one game, tend to behave similarly in another game <u>even if that behavior is neither efficient nor an equilibrium.</u> For example, the creation of a cooperative culture may be advanced by creating multiple institutions that create strong incentives for cooperation, so that cooperative behavior can then spill over into other contexts. This insight—that the ensemble matters—suggests that when we consider the performance of an institution, we should broaden our interpretive lens to include behaviors as well as outcomes, that we should not see behaviors as mere handmaidens of equilibria. Institutions induce behaviors that accumulate within individuals forming what we call behavioral repertoires. These repertoires become a part of what is more broadly defined as institutional context, be it organizational, ethnic, or national, and they, in turn, may well influence outcomes in other contexts.

Appendix. Experimental Instructions

We present the instructions for the (PD, WA) ensemble. Instructions for other ensemble treatments are identical except for the specific game forms. Instructions for the control sessions are standard. They are identical to the ensemble instructions except that two games and two other participants are replaced with one game and one other participant everywhere. Hence we omit them here.

Name: PCLAB: Total Payoff:

Introduction

- You are about to participate in a decision process in which you will play two games with two other participants. Each game will be played with a different participant and will be played for many rounds. This is part of a study intended to provide insight into certain features of decision processes. If you follow the instructions carefully and make good decisions, you may earn a considerable amount of money. You will be paid in cash at the end of the experiment.
- During the experiment, we ask that you please do not talk to each other. If you have a question, please raise your hand and an experimenter will assist you.

Procedure

- Matching: At the beginning of the experiment, you will be matched randomly with two other participants, both of whom will be your matches for the entire experiment. You will be matched with these same two people in all rounds. You will play a different game with each of these people.
- Roles: Throughout the game, you will be designated as the “row” player and your matches will be the “column” players. You will be a “row” player in all rounds, and your matches will be “column” players in all rounds
- Actions: In each round, you and your two matches will simultaneously and independently make decisions in two different games. One is the left game and the other is the right game.

You will play the left game with one of your matches (Left Game Match) and play the right game with the other match (Right Game Match). In each game, the row player (you) will click either the Top (A) or the Bottom (B) button. The column player (your Left or Right Game Match) will choose either the Left (A) or Right (B) button. These choices determine which part of the matrix is relevant (Top Left, Top Right, Bottom Left, Bottom Right).

- Interdependence: A player's earnings depend on the decision made by the player and on the decision made by his or her two matches as shown in the matrix below. In each cell, the row player's payoff is shown in red and the column player's payoff is shown in blue.

		Column Player			Column Player	
		Left (A)	Right (B)		Left (A)	Right (B)
Row	Top (A)	7, 7	2,10	Top (A)	7, 7	4,11
Player	Bottom (B)	10,2	4,4	Bottom (B)	11,4	5, 5

For example, if the row player (you) chooses Top (A) and the column player (your left game match) chooses Right (B) in the left game, then the row player (you) will get 2 points, while the column player (your left game match) will get 10 points in this game. Meanwhile, if the row player (you) chooses Bottom (B) and the column player (your right game match) chooses right (B) in the right game, then the row player (you) will get 5 points, and the column player (your right game match) will also get 5 points in this game. So as the row player in both games, you will get 7 points in this round totally.

- Rounds: You will first play the two games for 200 rounds. After round 200, whether the games will continue to the next round depends on the "throw of a die" that is determined by the computer's random number generator. At the end of each round after round 200, with 90% chance, the games will continue to the next round. With 10% chance, the games stop.

- Earnings: Your earnings are determined by the choices that you and your two matches make in every round. Your total earning is the sum of your earnings in all rounds.

  The exchange rate is $1 for <u>100</u> points.

  You can round up your total earning to the next dollar. For example, if you earn $15.23, you can round it up to $16.

- History: In each round, your and your two matches' decisions in all previous rounds will be displayed in a history window.

We encourage you to earn as much money as you can. Do you have any questions?

Figure 1: Control: Behavior in Self Interest

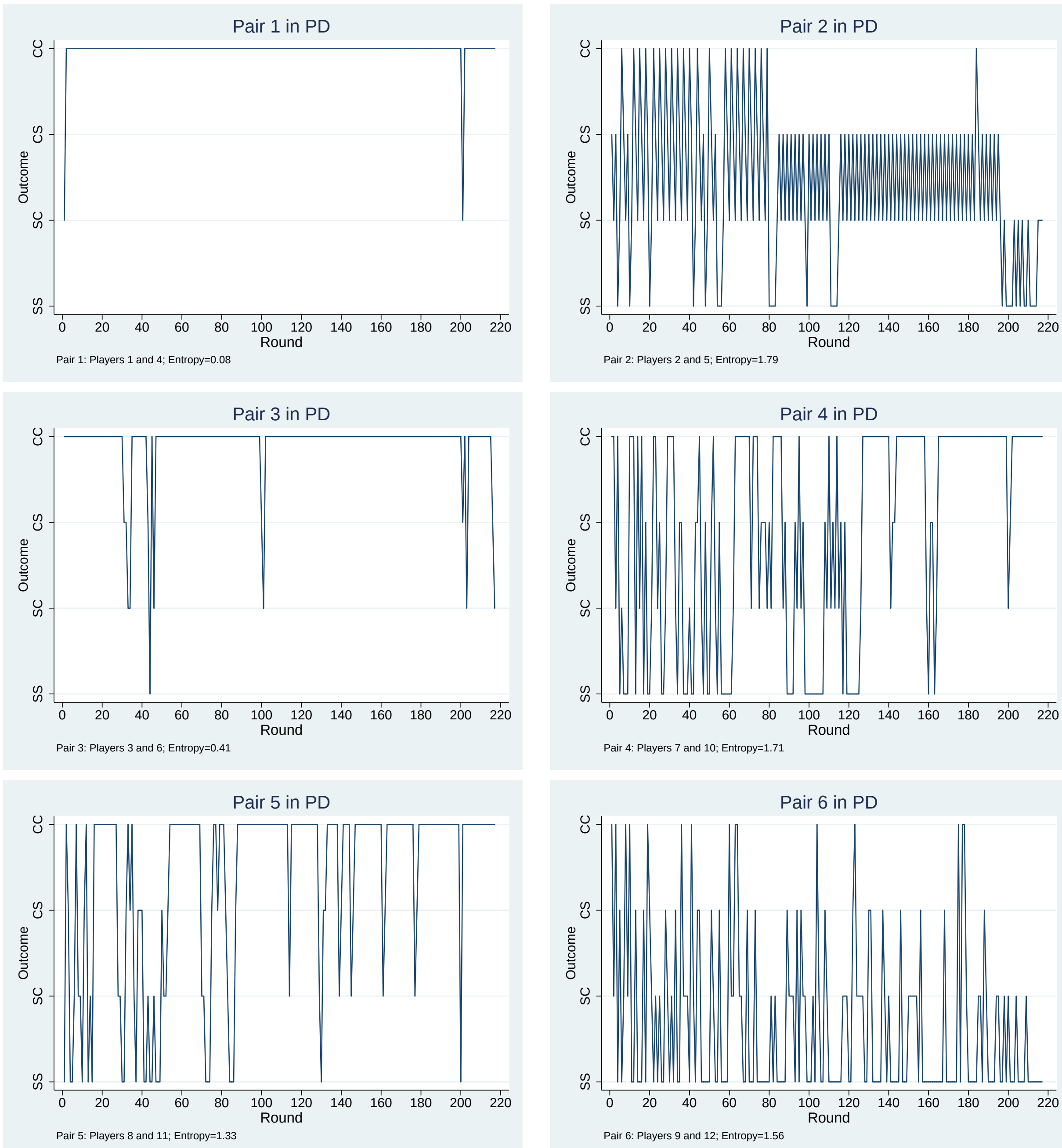


Figure 2: Control: Behavior in Prisoner's Dilemma

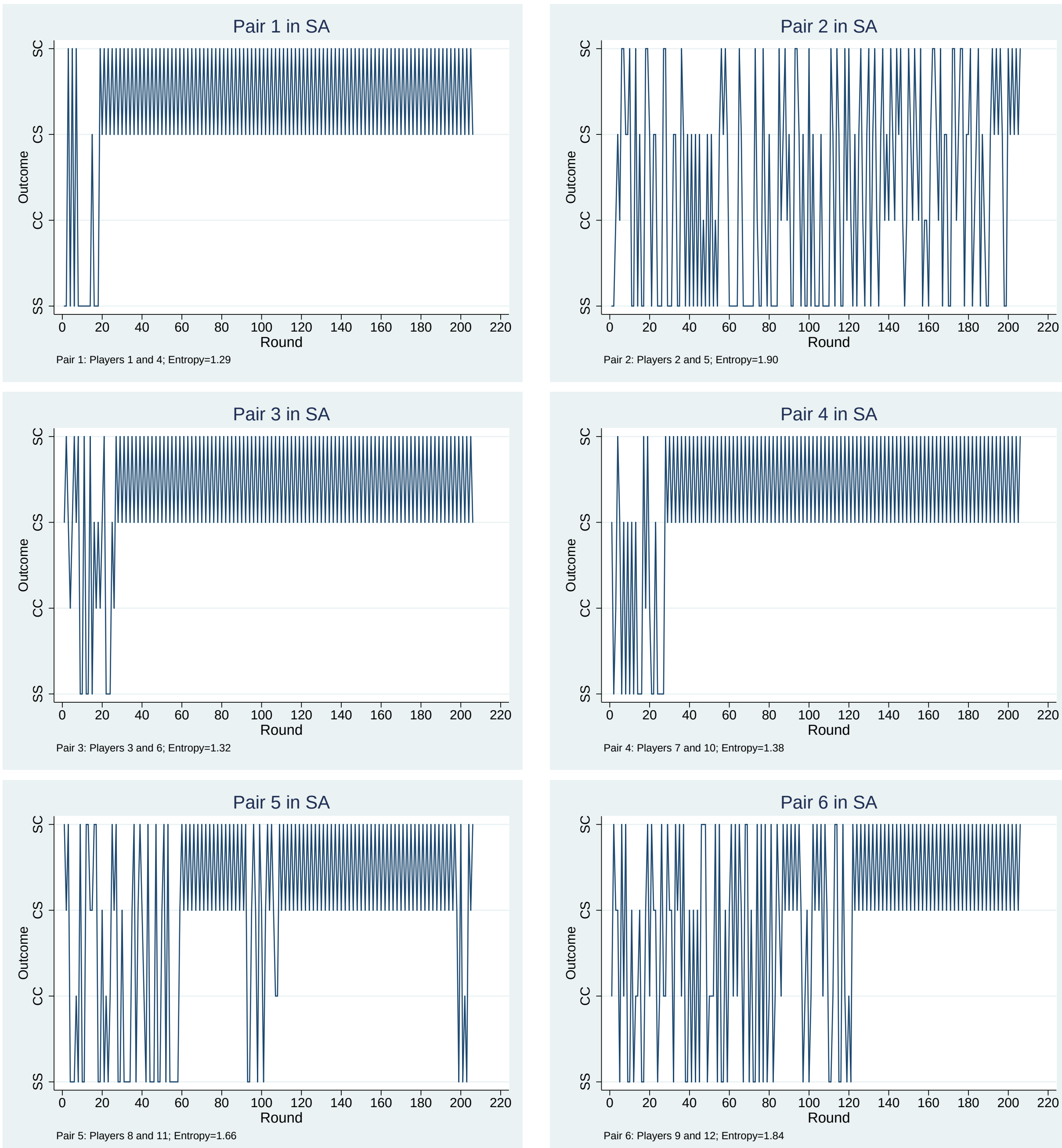


Figure 3: Control: Behavior in Strong Alternation

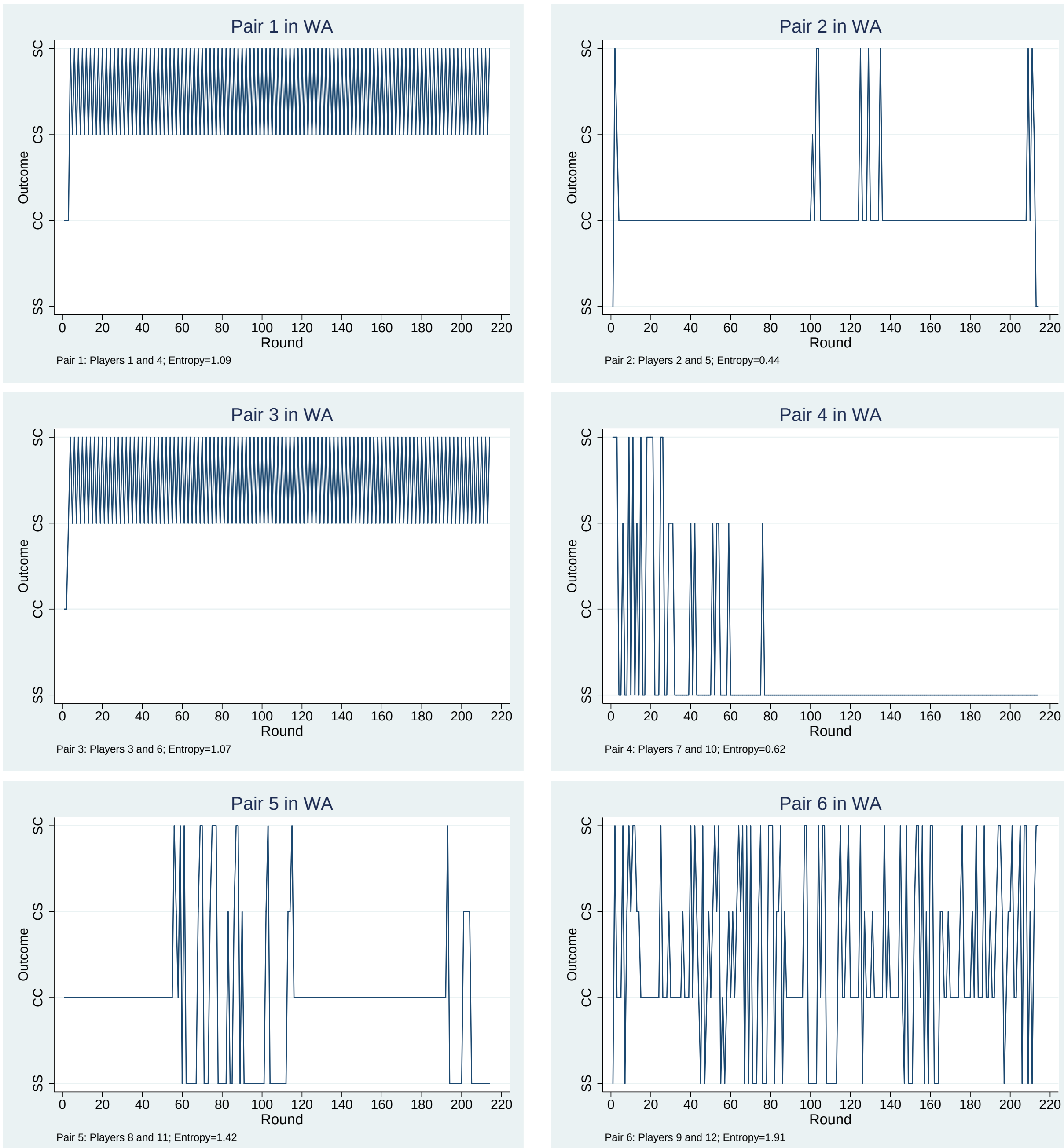
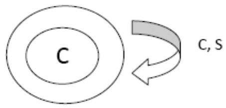
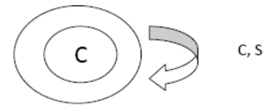


Figure 4: Control: Behavior in Weak Alternation

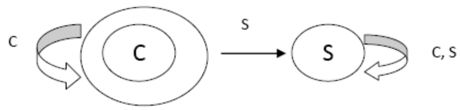
M1 (AC)



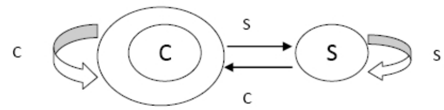
M2 (AS)



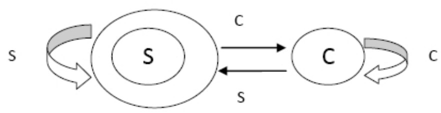
M3 (GT)



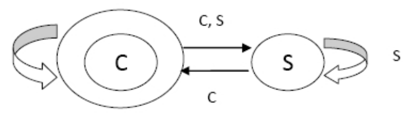
M4 (TFT)



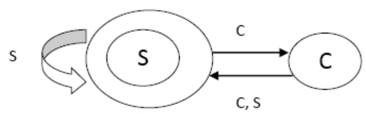
M5 (sTFT)



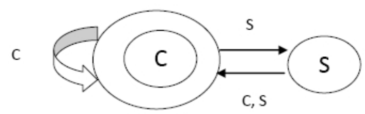
M6 (SAC)



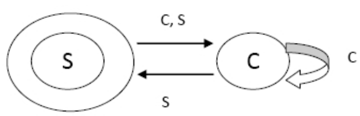
M7 (sSAC)



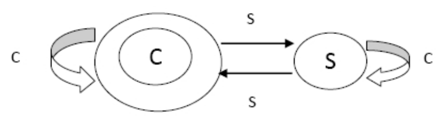
M8 (F1)



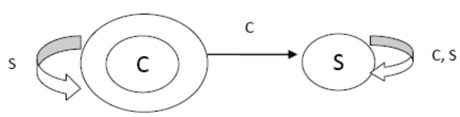
M9 (sF1)



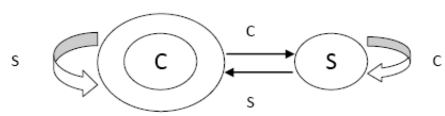
M10



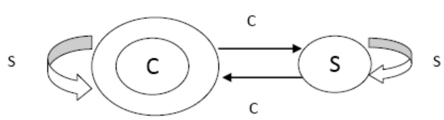
M11



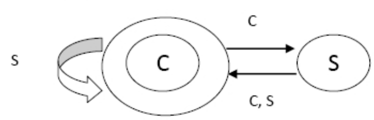
M12



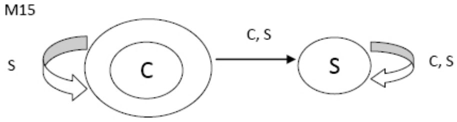
M13



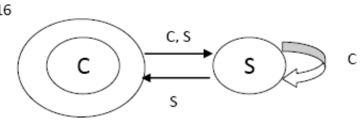
M14



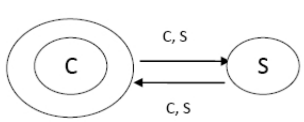
M15



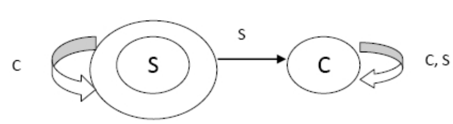
M16



M17



M18



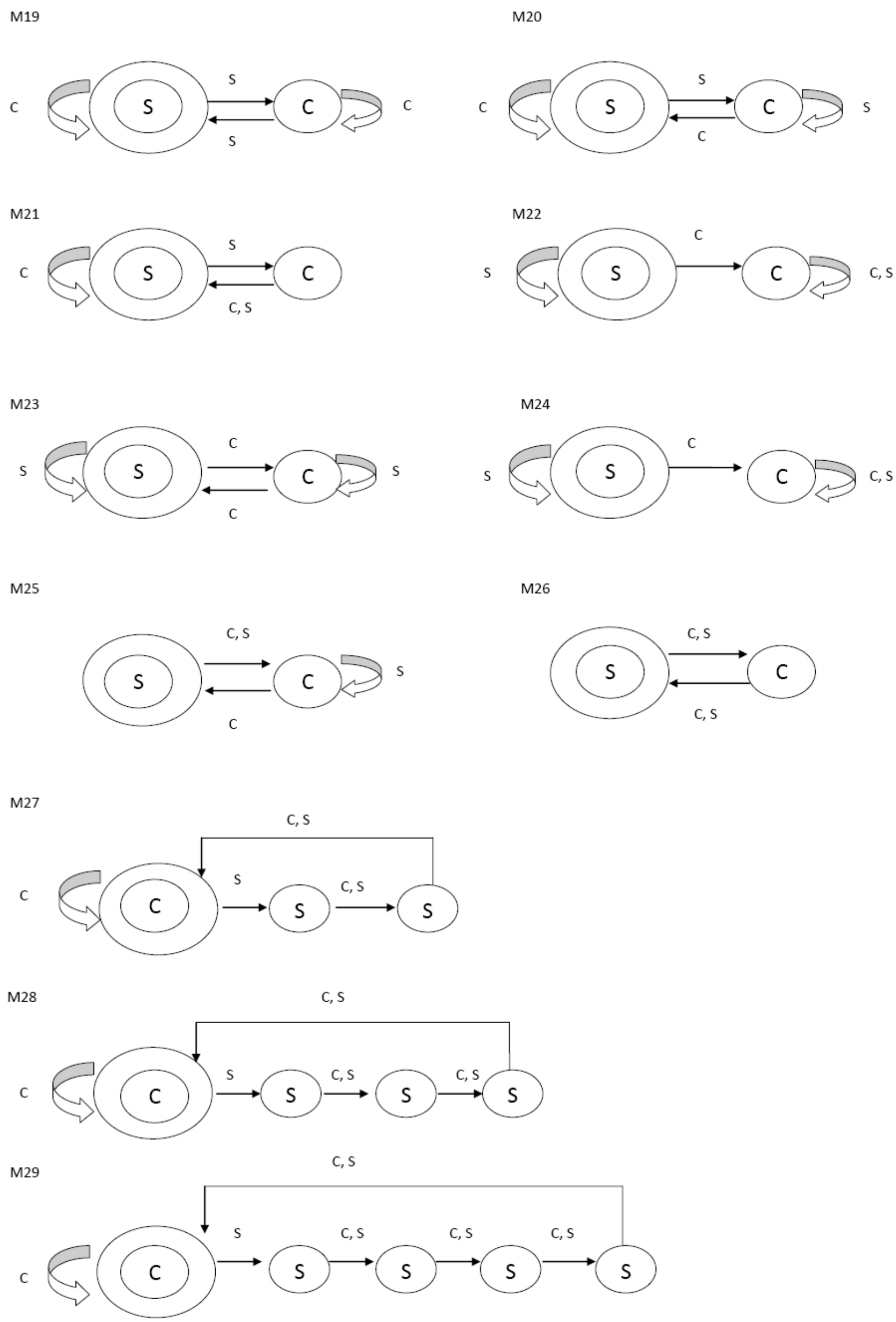


Figure 5: Automata Representation of Repeated Game Strategies

References

Abreu, Dilip and Ariel Rubinstein, "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica*, 1988, *56* (6), pp. 1259–1281.

Andreoni, James and John Miller, "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, 2002, *70* (2), 737–753.

Axelrod, Robert and William D. Hamilton, "The Evolution of Cooperation," *Science*, 1981, *211* (4489), 1390–1396.

Baron, David and Ehud Kalai, "The Simplest Equilibrium of a Majority-Rule Division Game," *Journal of Economic Theory*, 1993, *61*, 290–301.

Bednar, Jenna and Scott E Page, "Can Game(s) Theory Explain Culture?: The Emergence of Cultural Behavior in Multiple Games," *Rationality and Society*, 2007, *19* (1), 65–97.

Cason, Timothy N. and Lata Gangadharan, "Cooperation Spillovers and Price Competition in Experimental Markets," 2010. Working Paper.

__ , Anya Savikhin, and Roman M. Sheremeta, "Cooperation Spillovers in Coordination Games," 2010. Working Paper.

Cooper, David J. and John H. Kagel, "Learning and transfer in signaling games," *Economic Theory*, March 2008, *34* (3), 415 – 439.

Cover, Thomas M. and Joy A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2006.

Crawford, Vincent and Bruno Broseta, "What Price Coordination? The Efficiency-Enhancing Effect of Auctioning the Right to Play," *The American Economic Review*, 1998, *88* (1), 198–225.

Easterly, William, *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done SO Much Ill and SO Little Good*, New York: Penguin Press, 2006.

Engle-Warnick, Jim and Robert L. Slonim, "Inferring repeated-game strategies from actions: evidence from trust game experiments," *Economic Theory*, August 2006, *28* (3), 603–632.

Falk, Armin, Urs Fischbacher, and Simon Gaechter, "Living in Two Neighborhoods - Social Interaction Effects in the Lab," *Economic Inquiry*, 2010. Forthcoming.

Fischbacher, Urs, "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*, 2007, *10* (2), 171–178.

Gigerenzer, Gerd and Reinhard Selten, *Bounded Rationality: The Adaptive Toolbox*, Cambridge: MIT Press, 2002.

Gilboa, Itzhak and David Schmeidler, "Case-Based Decision Theory," *The Quarterly Journal of Economics*, 1995, *110* (3), 605–639.

Grief, Avner, *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*, New York: Cambridge University Press, 2006.

Grimm, Veronika and Friederike Mengel, "An Experiment on Learning in a Multiple Games Environment," 2010. Working Paper.

Ham, John C., John H. Kagel, and Steven F. Lehrer, "Randomization, endogeneity and laboratory experiments: the role of cash balances in private value auctions," *Journal of Econometrics*, 2005, *125* (1-2), 175 – 205. Experimental and non-experimental evaluation of economic policy and models.

Hanaki, Nobuyuki, Rajiv Sethi, Ido Erev, and Alexander Peterhansl, "Learning Strategies," *Journal of Economic Behavior and Organization*, 2005, *56*, 523–542.

Haruvy, Ernan and Dale O. Stahl, "Learning Transference Between Dissimilar Symmetric Normal-Form Games," 2010. Working Paper.

Huck, Steffen, Philippe Jehiel, and Tom Rutter, "Feedback Spillover and Analogy-based Expectations: A Multi-game Experiment," *Games and Economic Behavior*, 2010.

Jehiel, Philippe, "Analogy-based Expectation Equilibrium," *Journal of Economic Theory*, 2005, *123*, 81–104.

Kalai, E. and W.Stanford, "Finite Rationality and Strategic Complexity in Repeated Games," *Econometrica*, 1988, *56* (2), 397–410.

Miller, John H. and Scott E. Page, *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*, Princeton, NJ.: Princeton University Press, 2007.

North, Douglass C., *Understanding the Process of Economic Change*, Princeton, NJ: Princeton University Press, 2005.

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

Putnam, Robert D., *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton, NJ.:

Princeton University Press, 1993.

Rubinstein, Ariel, "Perfect equilibrium in a bargaining model," *Econometrica*, 1982, *50*, 97–109.

__ , "Finite automata play the repeated prisoner's dilemma," *Journal of Economic Theory*, 1986, *39* (1), 83 – 96.

Samuelson, Larry, "Analogies, Adaptations, and Anomalies," *Journal of Economic Theory*, 2001, *97* (2), 320–366.

Savikhin, Anya and Roman M. Sheremeta, "Simultaneous Decision-Making in Competitive and Cooperative Environments," 2010. Working Paper.

Shannon, Claude E., "A Mathematical Theory of Communication," *Bell System Technical Journal*, 1948, *27*, 379–423.

Tversky, Amos and Daniel Kahneman, "Rational Choice and the Framing of Decisions," *The Journal of Business*, 1986, *59* (s4), S251.

Van Huyck, John B., Raymond C. Battalio, and Richard O. Beil, "Asset Markets as an Equilibrium Selection Mechanism: Coordination Failure, Game Form Auctions, and Tacit Communication," *Games and Economic Behavior*, 1993, *5* (3), 485 – 504.

Weber, Roberto A., "Managing growth to achieve efficient coordination in large groups," *American Economic Review*, 2006, *96* (1), 114 – 126.