On the empirical relevance of correlated equilibrium

Daniel Friedman* Jean Paul Rabanal† Olga A. Rud‡
Shuchen Zhao§

September 29, 2021

Abstract

Can an efficient correlated equilibrium emerge without any exogenous benevolent agent providing coordinating signals? Theoretical work in adaptive dynamics suggests a positive answer, which we test in a laboratory experiment. In the well-known Chicken game, we observe time average play that is close to the asymmetric pure Nash equilibrium in some treatments, and in other treatments we observe collusive play. In a game resembling rock-paper-scissors or matching pennies, we observe time average play close to a correlated equilibrium that is more efficient than the unique Nash equilibrium. Estimates and simulations of adaptive dynamics capture much of the observed regularities.

Keywords: Correlated equilibrium, Laboratory experiment, Adaptive dynamics.

*Economics Departments, University of California Santa Cruz and University of Essex; dan@ucsc.edu

†Department of Economics and Finance, University of Stavanger, Norway; jeanpaulrab@gmail.com

‡Department of Economics and Finance, University of Stavanger, Norway; olga.rud@gmail.com

§Corresponding author; Institute for Advanced Economic Research, Dongbei University of Finance and Economics; fenix.zhaoshuchen@gmail.com; 350 W McKinley Ave Unit 515, Sunnyvale, California 94086, United States.

1 Introduction

A probability distribution φ over the set of action profiles in a normal form game is a *correlated equilibrium* (CE) if no player has an incentive to deviate from any action j in her component of the support of φ. As an equilibrium concept, CE has several advantages over Nash equilibrium (NE). First, as pointed out by Aumann (1974), the rationality assumptions are much simpler – common knowledge of rationality suffices for CE but not for NE. Second, CE is more general in that it does not require statistical independence — φ in mixed NE must be a player-by-player product distribution. Third, the set of all CE is convex and much easier to compute than the set of all NE, which may have disconnected components. Perhaps for these reasons, CE is the standard game theoretical equilibrium concept for many computer scientists.

Nevertheless, correlated equilibrium remains a curiosum for most economists and other social scientists. We believe that there are two reasons for this. First, the standard interpretation of CE is that the distribution φ is implemented via an exogenous benevolent agent who recommends a particular action to each player. Such agents, such as a traffic cop who cannot stop to write tickets, exist only in very special circumstances. Second, there are very few well known examples of games with a CE that achieves both efficiency and fairness between players but is not a NE.

The present paper attempts to surmount those obstacles by means of a laboratory experiment. We emphasize that the standard interpretation is unnecessarily restrictive. The definition of CE does not require any sort of traffic cop; it suffices that, whenever player i happens for whatever reason to choose a strategy j that is part of a CE, her beliefs do not discourage her from playing j, and those beliefs are not contradicted by her personal experience. We will see shortly that such beliefs can arise from certain sorts of adaptive processes. Therefore, our laboratory environment has no traffic cops but it offers a variety of opportunities to learn from personal experience in a variety of games (some better known than others) with an efficient and fair CE.

Our research questions include the following. In relevant games, does the observed time average joint distribution approximate a product distribution? If so, is it close to a NE distribution? If not, is it close to an efficient and fair CE distribution? Does the joint distribution display dynamic regularities?

In the following section we present an overview of the literature, including a handful of experiments studying CE; most of them rely on an external coordination device or some form of communication. Next, we present the theoretical foundations for our experiment. They derive largely from Foster and Vohra (1998), and from Hart and Mas-Colell (2000, 2001) who propose regret-based adaptive dynamics that promote convergence of the time average distribution φ to the CE set.

The Experimental Design section introduces the two games we study. The first is the well known Chicken game (CH), which has two NE in pure strategies as well as a mixed strategy NE and an attractive collusive strategy profile. Our second game is a variant of Matching Pennies proposed by Moulin and Vial (1978) (MV) that has no pure strategy NE nor an attractive collusive profile. The two games represent two typical circumstances where an CE shows an advantage over NE, thus generalize the results of our experiment.

Although we use a repeated game environment, our focus is stage game equilibria, consistent with our research questions. Fortunately, those research questions are best addressed in low information environments that are not conducive to well-known repeated game strategies such as grim trigger. In our low information environment, players do not know the actions or payoffs of other players, but they always know either their own historical average payoff for each action or, alternatively, the counterfactual payoff that they would have earned if they had always made a different choice.

The results are encouraging. We find in most treatments that aggregate joint distribution of actions is not well approximated by a product distribution. In CH games with low information, the aggregate joint distribution under historical average payoff is closer to a CE distribution than it is to any NE distribution, but with counterfactual payoff the distribution is consistently closer to an asymmetric pure NE. The disaggregated CH game results indicate collusive behavior in our higher information treatment. Our most interesting results are for the MV game, whose efficient and fair CE (unlike the unique mixed NE) puts zero weight on the cells with the lowest payoff for both players. We observe that in all of our treatments of the MV games, players indeed tend to avoid those low payoff cells.

The adaptive dynamics models lead us to estimate how players respond to "regret" given counterfactual or, alternatively, historical average payoff information. Using a logit model that allows for some inertia in the players' actions, we find that players indeed respond strongly to positive regret where they could have earned a higher payoff from an alternative action. Finally, we run simulations based on the logit estimates. The simulated outcomes track the laboratory outcomes fairly well and provide new insight into the observed regularities.

2 Previous Literature

Early literature on correlated equilibrium, first introduced by Aumann (1974, 1987), sought theoretical foundations for this equilibrium concept. For example, Brandenburger and Dekel (1987) show that common knowledge of rationality across players suffices to achieve CE, while Forges and Peck (1995) illustrate how the concept of CE can also represent a sunspot equilibrium in exchange markets. Moulin and Vial (1978) propose a new equilibrium concept called coarse correlated equilibria (CCE): the set of probability distributions over action profiles that can be supported if each player either commits to playing according to the recommendation of a device, or else plays freely with no access to the device recommendation. They note that the equilibrium sets satisfy NE $\subset$ CE $\subset$ CCE.

The theoretical literature on convergent dynamics is especially relevant. Fudenberg and Levine (1999) propose a smooth fictitious play procedure which can guarantee a almost sure convergence to the set of correlated ϵ-equilibria. Foster and Vohra (1997) introduce "calibrated" strategies with the property that time average counterfactual regret (defined in the next section) converges to zero irrespective what strategies the other players adopt. The authors show that time average play will converge to CE if all players adopt calibrated strategies. As explained in the next section, Hart and Mas-Colell (2000, 2001) introduce specific regret matching strategies that, if followed by all players, guarantee convergence to the set of CE. Metzger (2018) models evolutionary dynamics given exogenous coordinating signals, and finds that they can lead to CE outcomes that are not NE. Arifovic et al. (2019) implement an evolutionary learning simulations and obtain outcomes similar to those seen in the experiments noted below by Duffy and Feltovich (2010) and Duffy et al. (2017).

It is also worth noting that some learning rules which incorporate stochastic choice can also converge to a pure NE in two player games, using either sampling and regret computations (Foster and Young, 2006), or trial and error (Young, 2009). In the absence of a pure NE, Pradelski and Young (2012) propose that the long run behaviour under a log linear learning rule depends on the sum of payoff across all players, and the gain from a unilateral deviation by some player. In a low information public goods experiment, Nax et al. (2016) find that inertia, reversion and reinforced adjustment are key features of learning dynamics.

Thus far, empirical work on CE has been sparse and has focused mainly on laboratory experiments with exogenous coordinating signals. Chicken (or hawk-dove) games with private recommendations have been studied by Cason and Sharma (2007) and Duffy and Feltovich (2010). The former study concludes that recommendations are followed when subjects play automated counterparties who always follow recommendations, while the latter study finds that players follow recommendations when it implements a CE that is more efficient than relevant NE. Another game that has been studied in the laboratory is Battle of the Sexes (BoS). Duffy et al. (2017) find that direct messages improve coordination on a CE, relative to indirect messages. Anbarcı et al. (2018) find that subjects are more likely to follow recommendations when payoffs are more symmetric. Bone et al. (2013) find that subjects follow recommendations from a public device and coordinate more in a symmetric BoS than in the game of Chicken. Georgalos et al. (2020) consider a game proposed by Moulin and Vial (1978) that has a pure NE and a Pareto superior CCE. They find that a small and declining fraction of subjects choose the correlation device for that CCE; in key treatments most players prefer free choice, and thus their play eventually approximates the pure NE.

In an experiment without an external device but allowing for preplay communication, Moreno and Wooders (1998) study a constant sum, 3 player matching pennies game, and find that average play is a noisy version of a non-NE target CE. Palfrey and Pogorelskiy (2019) find support for CE in a voter turnout game with communication within parties. Furthermore, without an external correlation device, Cason et al. (2020) show that correlation of beliefs in some prisoner's dilemma games is less frequent than in hawk-dove (a variant of chicken) games and some coordination games.

The present empirical paper seems to be the first to focus on whether convergence to CE can arise from a regret-based adaptive process that requires no signal nor external devices nor communication among players.

3 Theoretical Considerations

Our experiment is shaped by existing theories. We begin with the standard definition: a probability distribution φ over the set $S = S^1 \times ... \times S^N$ of action profiles in an N player normal form game Γ is a *correlated equilibrium* (CE) if, for every player $i \in N$ and every two actions $j, k \in S^i$. we have

$$\sum_{s \in S: s^i = j} \varphi(s)[u^i(k, s^{-i}) - u^i(s)] \leq 0. \quad (1)$$

That is, no player has an incentive to deviate from any action j in her component of support φ. Nash equilibrium is the special case where φ is a product distribution, i.e., each player's realization from the mix is independent of other players' realizations. By contrast, in (1) the realizations can be correlated. The set of CE is characterized by the list of linear inequalities summarized by equation (1). That set is convex and can be computed using standard linear programming packages; see Appendix for an example.

3.1 Example Games

Table 1: Chicken (CH) game

	L	R
U	100, 100	600, 200
D	200, 600	500, 500

Table 1 shows a 2×2 bimatrix for the well-known Chicken Game (CH) with three possible NE: two asymmetric pure NE $(U, R), (D, L)$ and the symmetric mixed NE (MNE) $(.5, .5) = \frac{1}{4}$(U,L)+ $\frac{1}{4}$(D,L) + $\frac{1}{4}$(U,R) + $\frac{1}{4}$(D,R). It also has an efficient and fair CE, $\frac{1}{3}$(D,L) + $\frac{1}{3}$(U,R) + $\frac{1}{3}$(D,R), that we will refer to as our *target CE*. The expected payoff vector $(\frac{1300}{3}, \frac{1300}{3})$ of the target CE has higher efficiency (i.e., higher payoff sum) and fairness (i.e., smaller payoff difference) than any of the NE. Note that pure collusion (D,R) is the most efficient and fair profile, but it is not a CE (nor, a fortiori, a NE).

The set of mixed profiles for CH is a three dimensional subset of four-dimensional space. It may be more useful to consider the two-dimensional payoff space representation shown in Figure 1, where the payoff vectors associated with CE lie in the blue quadrilateral, which covers $\frac{4}{15} \approx 27\%$ of the feasible payoff region.[1]

Table 2: Moulin and Vial (1978) (MV) game

	L	C	R
T	0, 0	100, 200	200, 100
M	200, 100	0, 0	100, 200
B	100, 200	200, 100	0, 0

Table 2 presents a 3x3 game based on Moulin and Vial (1978). This game, denoted MV below, resembles the standard rock-paper-scissors game in that it has a

[1]The ratio of the areas is the fraction of the main diagonal in the blue region. The main diagonal has length $5\sqrt{2}$ and its blue segment has length $(\frac{13}{3} - 3)\sqrt{2}$.

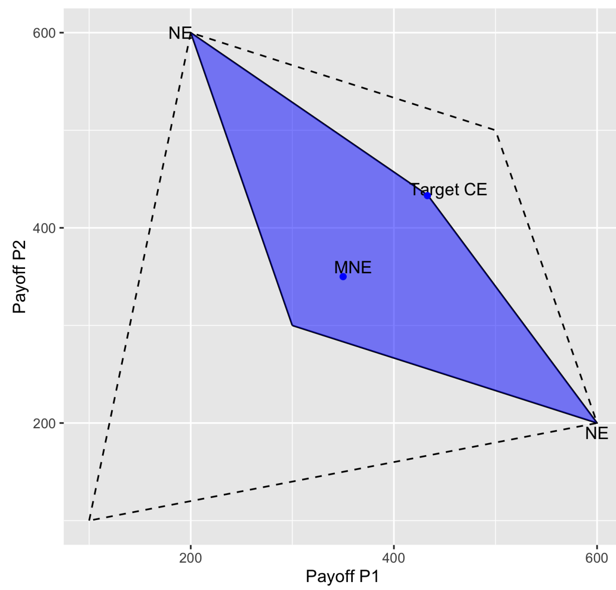


Figure 1: Payoff space for Chicken game (CH). The blue shaded region represents payoff vectors for all CE profiles, while payoffs for all feasible profiles are bounded by the dashed lines. The mixed NE payoff vector (350,350) is Pareto dominated by target CE payoff ($\frac{1300}{3}$, $\frac{1300}{3}$), which is dominated by collusion payoff (500,500). NE denotes the asymmetric pure NE payoffs, e.g., (200,600), whose sum is less than that of the target CE.

best response cycle. It is also reminiscent of matching pennies in that the off-diagonal profiles are constant sum, and therefore provide little scope for collusion. The MV game has a unique NE: the symmetric uniform independent mixture $(1/3, 1/3, 1/3) \times (1/3, 1/3, 1/3)$. This mixed NE is Pareto dominated by the *target CE*, which assigns a probability of zero to the main diagonal profiles and a probability of $\frac{1}{6}$ to the six other profiles. Figure 2 shows the feasible payoff space, and the CE region which covers $1/3^2 \approx 11\%$ of the feasible space.

3.2 Regret

We now present adaptive dynamics models specifying how players compare current action j to alternative actions k and how they use those comparisons to choose next period's action. These models closely follow Hart and Mas-Colell (2000), and require only that players know the history of realized profiles so far and their own payoff function. They do not require the knowledge of other players' payoffs, nor an exogenous agent to provide coordinating signals. Instead, correlation comes from a common

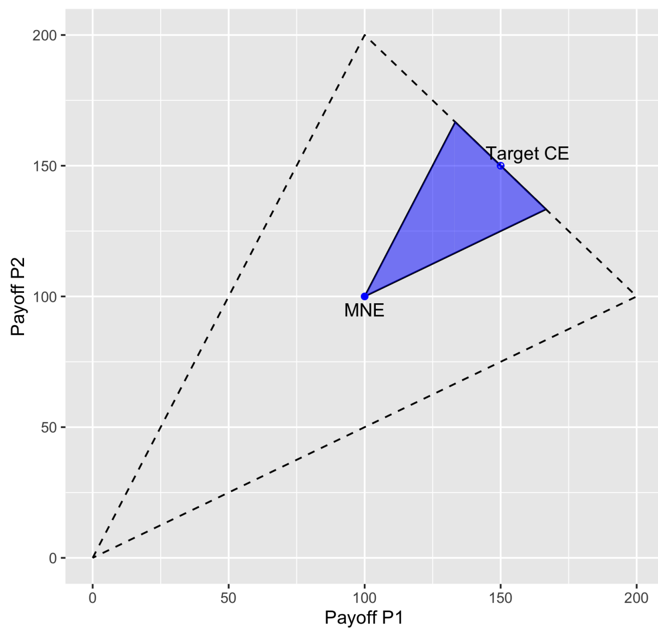


Figure 2: Payoff space for MV game. The blue region represents CE payoff vectors, while the dashed boundary encloses all feasible payoff vectors. The MNE payoff (100,100) is Pareto dominated by the target CE payoff (150,150).

history of play.

To formalize, suppose that the normal form game Γ is played repeatedly in discrete time $t = 1, 2, \ldots, T \leq \infty$. Let $s_t^i \in S^i$ denote the (realized) choice of player i at time t, and let $u^i(s_t)$ denote the payoff of player i in period t. For any two distinct actions $j \neq k \in S^i$ for any given player i at any time $t \leq T$, suppose that player i had replaced action j, every time $\tau \leq t$ that it was played so far, by action k, with no other changes in the profiles. Then player i's payoff at time τ becomes $u^i(k, s_\tau^{-i})$ if $s_\tau^i = j$. **Signed regret** is resulting difference in i's per period payoff so far,

$$\hat{r}_t^i(j,k) = \frac{1}{t} \sum_{\tau \leq t : s_\tau^i = j} \left[u^i(k, s_\tau^{-i}) - u^i(s_\tau) \right]. \tag{2}$$

The intuition is that a player may wish that they had done something different to the extent that it could have generated a higher payoff in the past, assuming no other changes on the part of other players.

Hart and Mas-Colell (2000) focus on what we refer to as **truncated regret**,

$$\hat{r}_t^i(j,k)_+ = \max\{\hat{r}_t^i(j,k), 0\}. \tag{3}$$

The intuition for this specification is that when $\hat{r}$ is negative, the magnitude or extent of signed regret is not important – the player is happy with the current action j is not tempted to try the alternative action k.

The focus of our analysis is **counterfactual regret**, a re-normalization of signed regret which looks only at periods where the counterfactual is relevant,

$$r_t^i(j,k) = \frac{\sum_{\tau \le t : s_\tau^i = j} u^i(k, s_\tau^{-i})}{|\tau \le t : s_\tau^i = j|} - \frac{\sum_{\tau \le t : s_\tau^i = j} u^i(s_\tau)}{|\tau \le t : s_\tau^i = j|} \equiv m_t^i(j,k) - M_t^i(j). \tag{4}$$

The last expression in equation (4) summarizes counterfactual regret as the signed difference between $m_t^i(j,k)$, which is the per-period mean counterfactual payoff of playing k instead of j whenever j was actually played, and $M_t^i(j)$, the mean payoff so far from playing j. Such normalization (averaging payoffs only for periods when j is played) is useful because, for any strategy j that is rarely played, the unnormalized signed regrets (or truncated regrets) all will automatically become small as t gets large, regardless of whether the alternatives j have relatively high or low payoff.

Similarly, we can write

$$r_t^i(j,k)_+ = \max\{r_t^i(j,k), 0\}. \tag{5}$$

For comparative purposes, we also consider **average regret**, defined as the difference in per period actual payoffs between an alternative action and the current action,

$$R_t^i(j,k) = \frac{\sum_{\tau \le t : s^i = k} u^i(s_\tau)}{|\tau \le t : s^i = k|} - \frac{\sum_{\tau \le t : s^i = j} u^i(s_\tau)}{|\tau \le t : s^i = j|} \equiv M_t^i(k) - M_t^i(j). \tag{6}$$

One could also write out truncated or unnormalized versions of average regret, but we will not use them in our experiment. The main interest in average regret comes from its informational economy: players can use it even when they don't know their payoff function, and only know their own action and payoff history.

How do the players respond to regret? According to Hart and Mas-Colell (2000), the action of player i in period $t+1$ is then chosen via a linear probability model using an appropriate version of regret r. The **HM response** rule states that given current action j, a player will choose action k next period with probability

$$p^i_{t+1}(k) = \frac{1}{\mu} r^i_t(j,k)_+, \quad k \neq j \qquad (7)$$
$$p^i_{t+1}(j) = 1 - \sum_{k \in S^i : k \neq j} p^i_{t+1}(k),$$

where the parameter $\mu > 0$ is large enough to ensure that $p^i_{t+1}(j)$ is positive, i.e., that the player will always continue to play strategy j with a positive probability. A larger μ indicates greater inertia. In continuous time, the degree of inertia can be tied to a sampling interval $\Delta t < 1$ by changing the scaling factor from $\frac{1}{\mu}$ to $\frac{\Delta t}{\mu}$.

An alternative response rule, widely used in empirical literature, is that choices follow the logit model. The **logit response** rule with parameter $\beta > 0$ sets

$$p^i_{t+1}(k) = \frac{e^{\beta r^i_t(j,k)}}{\sum_{\ell \in S^i} e^{\beta r^i_t(j,\ell)}} = \frac{e^{\beta m^i_t(j,k)}}{\sum_{\ell \in S^i} e^{\beta m^i_t(j,\ell)}}, \quad \text{for all } k \in S^i. \qquad (8)$$

The last expression uses equation (4) to cancel the common factor $e^{-\beta M^i_t(j)}$ in numerator and denominator.

Compared to (7), equation (8) assigns no special inertia to the current action, even though that might matter in practice. Therefore we also consider the **inertial logit response** rule with parameters $\beta > 0$ and $\Delta t \in (0, 1)$, defined as

$$p^i_{t+1}(k) = \frac{e^{\beta r^i_t(j,k)} \Delta t}{\sum_{\ell \in S^i} e^{\beta r^i_t(j,\ell)}}, \quad k \neq j \qquad (9)$$
$$p^i_{t+1}(j) = 1 - \sum_{k \in S^i : k \neq j} p^i_{t+1}(k),$$

where j denotes the action of player i at time t. When Δt is small, the player is more likely to stay with the current strategy j; when the player switches to a different action, the relative probabilities of alternative actions are governed by logit choice.

3.3 Hart and Mas-Colell convergence results

Hart and Mas-Colell (2000) Main Theorem: If every player follows the response rule (7) applied to (unnormalized counterfactual) truncated regret (3), then the time average profile $z_t = \frac{1}{t} \sum_{\tau \leq t} s_\tau$ converges almost surely to the set CE of correlated equilibria as $t \to \infty$.

Recall that truncated regret, like other versions of counterfactual regret, implicitly assumes that each player i knows the complete profile history and her own payoff function. Hart and Mas-Colell's follow-up paper (Hart and Mas-Colell, 2001) tweaks unnormalized truncated average regret via diminishing trembles and obtains convergence to CE even when players only know their own action and payoff history.

Note that these conclusions concern convergence to the CE set, and not to any particular point within the set. Hart and Mas Colell's papers offer few hints about whether a regret-based adaptive process will converge to a particular CE of interest.

4 Experimental Design

We run a full factorial design with three factors (or treatment variables). The first treatment variable is the game that subjects play — either standard Chicken (CH) as shown in Table 1 or Moulin and Vial (1978) (MV) as shown in Table 2. Recall that the NE sets and collusion possibilities differ sharply between these two bimatrix games.

The second treatment variable is the level of information provided by the user interface. In low information treatment (L), subjects see very little information besides that required to verify the regret display. High information treatment (H) includes information on payoff functions and opponent actions, and thus brings us closer to laboratory environments used in most previous bimatrix game experiments. The point is to check the robustness of behavior to additional feedback beyond the regret information.

The third treatment variable is the type of regret subjects see in the user interface. The components of regret are displayed as orange bars next to each radio button, as shown in Figure 3. In Counterfactual (**C**) treatments, these bars display the counterfactual per period payoffs $m_t^i(j,k)$ defined in equation (4); recall that $m_t^i(j,j) = M_t^i(j)$ is the actual historical average payoff for the current action. In the alternative Average (A) treatments, the orange bar for each action k shows its actual historical per-period payoff $M_t^i(k)$.[2]

Figure 3 shows an example user interface for treatment CH×L×{C, A}; as noted earlier, the orange regret bars are computed differently for C than for A treatments. The green bar in the upper right corner shows the time remaining until the player's action (here, either A or B, entered by clicking a radio button) becomes final for that

[2]We use normalized regret because otherwise the orange bars could shrink to invisibility in the later periods. One could, of course, rescale them from time to time, but that would seem to be more erratic than our normalization.

Figure 3: User interface for low information (L) Chicken (CH) game. The horizontal orange bars next to the radio buttons A and B (with text 380 and 350) show the current regret components. The upper graph shows the payoffs earned each period so far, while the lower graph shows the corresponding actions selected. The green bar in top right corner shows time remaining in the current period.

period. There are two time graphs which show current and past payoffs, and current and past actions, for each period so far.

Figure 4 shows the screen for the high information treatments CH×H×{C, A}. Beyond the information conveyed in L treatments, subjects in H treatments can observe (i) current and past actions, and payoffs, of the other player (P1 in the figure), (ii) their own payoff matrix, and (iii) the frequency of past profiles. The latter is shown by the intensity of purple shading across profile cells — the more intense the purple color, the more frequently this cell has been played. Lastly, the filled circles indicate the player's own current choice and last period's choice by P1. User interfaces for the MV games are analogous.

4.1 Procedures

Treatment variables are held constant within each session and varied across sessions. Each session consists of 2 practice supergames with 20 periods each, and 8 salient supergames with 50 periods each. Across all L treatments, each period within each supergame lasts 4 seconds. In H treatments, periods are 8 seconds for the first 4 supergames, and then shorten to 6 seconds for the last 4 supergames. The additional time in H treatments allows subjects to absorb the greater amount of information in the

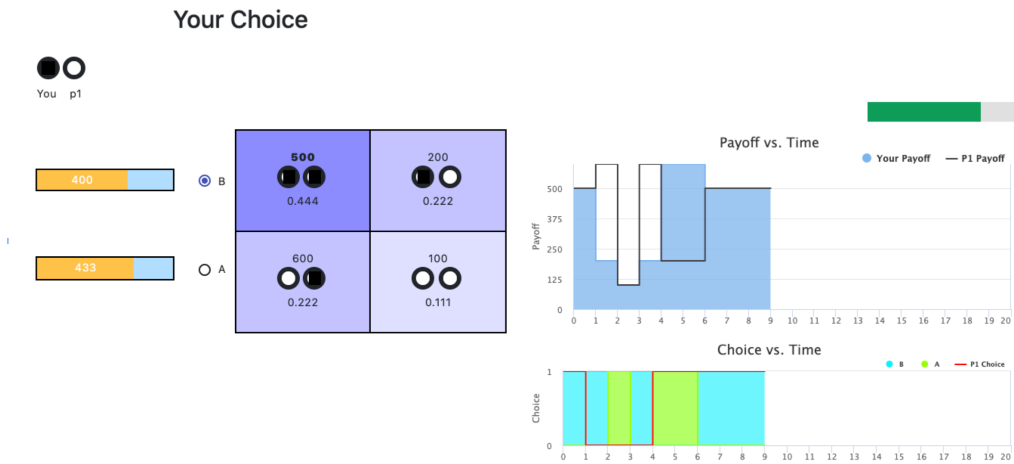


Figure 4: User interface for high information (H) Chicken (CH) game. Horizontal bars next to radio buttons still represent current regret components (here 433 and 400). Each cell of the 2×2 matrix reports own payoff (top number), own previous choice (first filled or empty circle), counterparty's previous choice (second circle), and the frequency of past play (bottom number, and shading intensity). The upper graph shows own (blue area) and counterparty payoffs (black line) so far. The bottom graph displays own choices (color coded), and counterparty (binary coded red line) so far.

display. The matching of players was random (fixed) between (within) supergames. We change the order of actions in the payoff matrices every other supergame to encourage subjects to remain attentive.

Table 3: Session information.

Game	Information	Regret	#subjects	#sessions
CH	L	Average	44	4
CH	H	Average	22	2
MV	L	Average	50	4
MV	H	Average	24	2
CH	L	Counterfactual	44	4
CH	H	Counterfactual	26	2
MV	L	Counterfactual	42	4
MV	H	Counterfactual	22	2

Table 3 summarizes the sessions by treatment. We completed 4 low information sessions and 2 high information sessions for each type of game and each type of regret. There were 10 to 14 participants in each session, with a total of 274 subjects in the 24 online sessions using Zoom. Subjects were drawn from the MonLEE (Monash University) subject pool. The experimenter read the instructions aloud, and answered questions via private chat. Participants received all points earned in the 8 salient supergames, which were converted to Australian Dollars (AUD) at the rate of 0.63 per 100 points in CH sessions, and 1.89 per 100 points in MV sessions. On average, subjects received 19 AUD (treatment averages ranged from 17.25 to 19.92) on top of the 5 AUD show-up fee. All payments were made via bank transfer. Sessions in low information treatments lasted less than one hour, and lasted up to 1.5 hours in high information treatments.

4.2 Hypotheses

Hypothesis 1. *In later periods, the overall observed (row, column) joint distribution of choices will be correlated. That correlation will be higher under (a) counterfactual regret than average regret, and (b) low information than high information.*

Hypothesis 1 focuses on aggregate time average profile frequencies and would fail if player pairs consistently approximate any specific NE or if, for any other reason, they choose their actions independently. Comparing Hart and Mas-Colell (2000, 2001) suggests that counterfactual regret is more conducive than average regret to CE convergence, and that in turn suggests higher correlation as in H1(a). Hypothesis H1(b) arises from the conjecture that the additional information provided in H treatments may distract attention from the regret bars.

Hypothesis 2. *Time average payoff vectors for player pairs will disproportionately lie in the CE region. The fraction of vectors in the CE region will be higher under (a) counterfactual regret than average regret, and (b) low information than high information.*

Hypothesis 2 applies at the level of individual pairs, and in that sense pertains more directly to CE play and addresses across-pair heterogeneity. Theoretical results of Hart and Mas-Colell and others do not predict convergence to a particular point in the CE set, so H2 refers to the entire CE region of payoff space. On the other hand, it would be especially interesting if convergence were to a Pareto efficient and "fair" (equal payoffs for the two players) CE in our symmetric games. Therefore we will also look for convergence to the target CE payoff vector. As before, we expect low information to make counterfactual regret especially salient and so facilitate convergence to the CE region.

Hypothesis 3. *Subjects will more likely switch to an alternative action the larger its positive regret. Response to negative regret will be weaker.*

To test this Hypothesis, we will estimate individual behavior in logit and HM switching models. H3 implies (a) positive coefficient estimates for positive regret and (b) less positive (perhaps zero, as per Hart and Mas-Colell) estimates for negative regret.

5 Results

Our main interest is in long-run behavior in player pairs, and whether frequencies (time averages) converge to a CE distribution. We are less interested here in transient learning behavior. Therefore, except when otherwise noted, the data reported below include only the last 60% of observations from each supergame (last 30 of 50 periods).

5.1 Aggregate behavior

Table 4 presents aggregate time average profile frequencies by treatment across CH games. The left panel summarizes the frequency of play in low information treatments, for Average (A) as well as for Counterfactual (C) regret, while the right panel does the same for high information treatments. Recall that the target CE puts frequency zero in the UL cell and 0.33 in the other cells. The Table shows that this target CE is best approximated with Average regret and Low information: the lowest frequency of play here is indeed the UL cell (0.12), while the collusion cell DR has frequency 0.39, modestly higher than the target 0.33. With Counterfactual regret, subjects mostly play the two pure NE cells (UR and DL) with similar frequencies (0.40 and 0.45), and put little weight on the main diagonal (0.07 each on UL and DR). In the high information treatment, collusive behavior is dominant for both Average (frequency 0.44) and Counterfactual (0.58) regret.

Table 4: Time average frequency in CH games

	Low information					**High information**			
	A		**C**			**A**		**C**	
	L	R	L	R		L	R	L	R
U	0.12	0.22	0.07	0.40	U	0.23	0.18	0.17	0.12
D	0.27	0.39	0.45	0.07	D	0.16	0.44	0.14	0.58

Table 5 presents time average frequencies for sessions using the MV game. Here all treatments yield roughly similar outcomes; consistent with the target CE for this game (but not its unique NE), all diagonal cells have lower frequency than the other six off-diagonal cells. This is especially true in the C treatments.

Table 5: Time average frequency in MV games

	Low information							**High information**					
	A			**C**				**A**			**C**		
	L	C	R	L	C	R		L	C	R	L	C	R
T	0.07	0.14	0.13	0.04	0.14	0.12	T	0.06	0.15	0.12	0.04	0.13	0.10
M	0.13	0.07	0.13	0.13	0.05	0.12	M	0.11	0.08	0.13	0.15	0.06	0.17
D	0.12	0.15	0.07	0.18	0.17	0.06	D	0.14	0.12	0.08	0.15	0.15	0.06

Thus the evidence so far qualitatively supports

Result 1. *Consistent with Hypothesis 1, time average action profiles in later periods are not well approximated by row × column product distributions. Consistent with H1(a), the approximation is worse under counterfactual regret than average regret.*

To test Hypothesis 1 more quantitatively, we adapt a maximum likelihood test introduced by Moreno and Wooders (1998). The null hypothesis states that the log-likelihood is generated by a product distribution. Using the time average relative frequencies reported in Table 4 for the CH game, and in Table 5 for MV game, we compute the number of independent observations needed to reject the null at critical value $\chi^2_{0.05}$ with one degree of freedom for CH and four degrees of freedom for MV. Those numbers of independent observations indicate the difficulty of rejecting the null hypothesis given our aggregate data.

For the CH game, we find that only 7 such observations would be required given the observed cell proportions in L×C. For the H×C data, 26 such observations would be required, and 960 for L×A and 41 for H×A. For the MV game, the minimum number of independent observations required for L×C and H×C is similar (about 60), while the numbers for L×A and H×A are higher (118 and 155, respectively). Note that the numbers of independent observations required are much smaller than the number of actual (but not necessarily independent) observations in our sample.[3] Thus, with the possible exception of L×A in the CH game, we can clearly reject the null hypothesis of no correlation across player actions.

[3]Table 6 reports the number of pairs for each treatment combination, which should be multiplied by 30 to obtain the actual number of observations in the sample.

Table 6: Number of observations required for rejection in the LR test

	CH		MV	
	L	H	L	H
A	960 (133)	41 (74)	118 (170)	155 (83)
C	7 (140)	26 (90)	60 (140)	59 (74)

Notes: The LR test is adopted from Moreno and Wooders (1998). In parentheses is the number of pairs observed in each treatment combination. Since each pair is observed over 30 periods, the number of observations in our sample is 30 times larger.

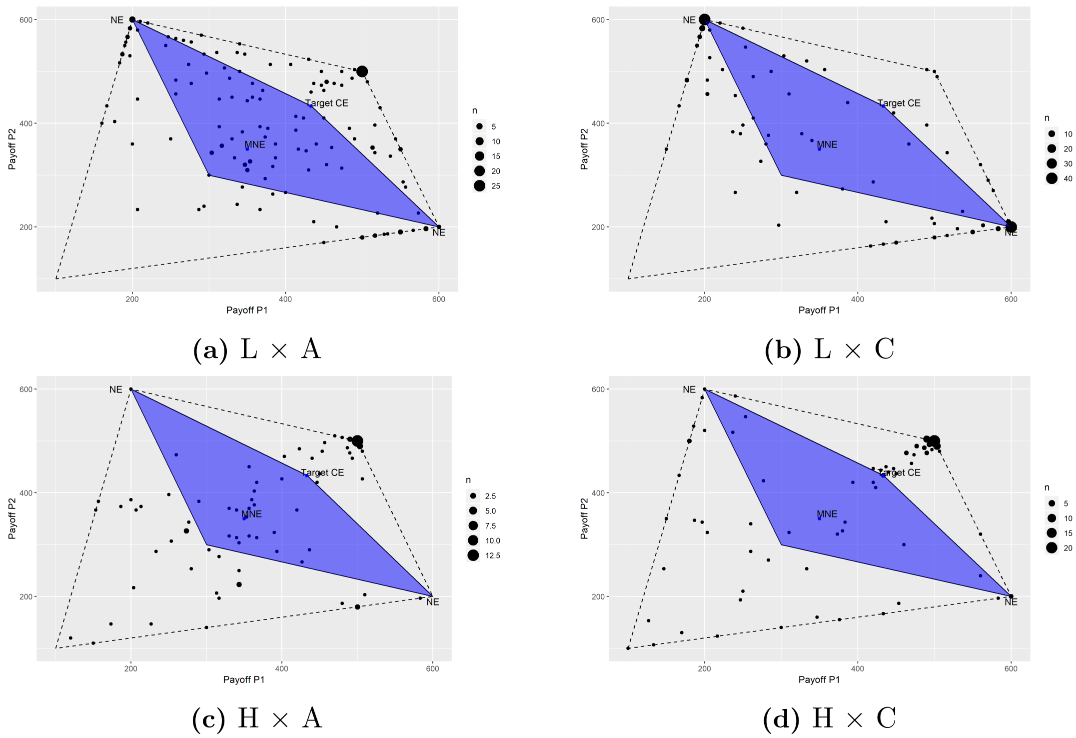
We can also conclude that the approximation of time average profiles by a product distribution is worse under counterfactual regret than average regret, consistent with H1(a). On the other hand, the evidence the H vs L treatment effect is mixed. Consistent with H1(b), in CH×C sessions the null hypothesis of product distribution is rejected faster under low information, while the opposite holds in CH×A sessions. We see no information treatment effect in MV×C sessions, and perhaps weakly favorable evidence for H1(b) in MV×A sessions.

5.2 Player pair average behavior

Figure 5 shows pair-level results relevant to Hypothesis 2 from Chicken (CH) games. In sessions with treatment CH×L×A, the Figure shows that many pairs collude to obtain payoffs (500, 500); vectors for other pairs are mostly sprinkled in or near the CE region. For CH×L×C sessions, most pairs play one of the two pure NE, whose payoff vectors (200, 600) and (600, 200) are corner points of the blue CE quadrilateral. For both regret conditions in the high information treatments, a large fraction of pairs display a high degree of collusion. In CH×H×A sessions, many pairs that do not collude play close to the MNE. In corresponding C sessions there are more pair payoff vectors close to the target CE, and between it and the collusion payoff vector.

By Figure 5 we also conclude that H information treatment has a great impact on the subjects' behavior with counterfactual regret, but not with average regret. The counterfactual regret leads the subjects to stay at either pure NE. However, with more information, the subjects are much more likely to figure out the pattern of the game and collude to improve fairness and social welfare. The average regret does not necessary lead the subjects to stay at NE, thus does not show a large difference between the two information treatments.

Figure 6 similarly summarizes the pair payoffs for the four MV treatments. The



(a) L × A **(b)** L × C **(c)** H × A **(d)** H × C

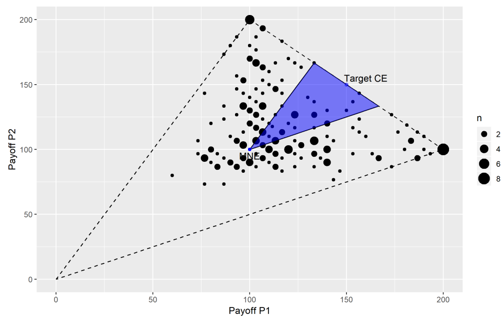
Figure 5: Payoffs for player pairs in CH games. The four panels separate data by combinations of L vs H information and A vs C regret. The CE region is shaded blue.

time average payoffs are mostly in or near the CE zone, especially in L×C sessions. Payoff vectors tend to be closer to the target CE (and are more likely to be northeast of the MNE) in Counterfactual than in corresponding Average regret treatments.

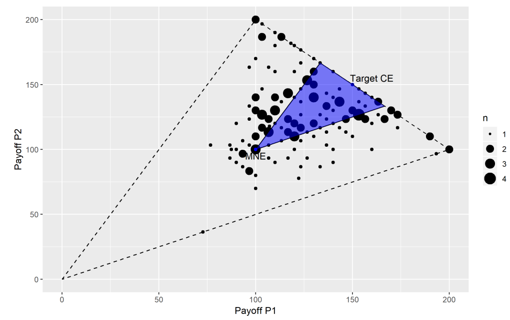
Thus the evidence so far qualitatively supports

Result 2a. *Consistent with Hypothesis 2, player pairs in CH games with Low information are disproportionately likely to earn average payoffs in the CE region. The same is true for all treatments using MV games. Consistent with H2(a,b), for each game the fractions of pair payoffs in the CE region are largest in L×C treatments.*

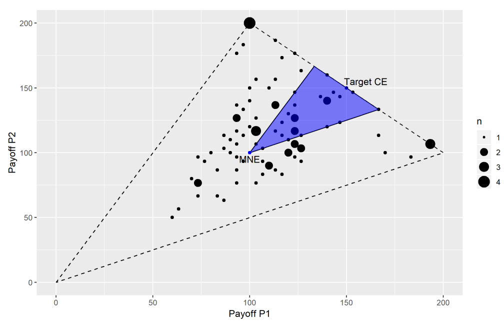
Table 7 offers a more quantitative test of H2. The first column presents for each treatment the observed fractions of pair payoffs that are in the CE region. We perform a proportional test with null hypothesis that the observed fraction of pairs is no greater than the uniform random success rate of 0.27 for the CH game and 0.11 for the MV game; recall from Section 3.1 that these rates represent the fractions of feasible payoff space that lie in the CE region. For the CH game, we fail to reject that the observed fraction is equal or lower to a random draw for H×A and H×C. As claimed in Result 2a,



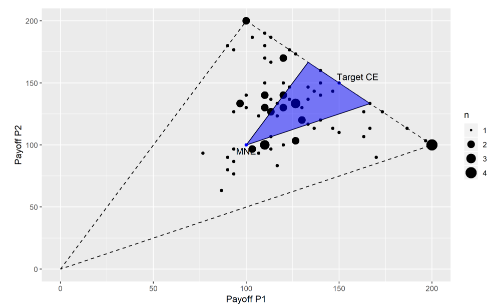
(a) L×A



(b) L×C



(c) H×A



(d) H×C

Figure 6: Payoffs for player pairs in CH games. The four panels separate data by combinations of L vs H information and A vs C regret. The CE region is shaded blue.

Table 7: Fraction of pairs in CE region; distance to target CE

	Fraction in CE	Distance Target CE
CH game	0.27	
L×C	0.57 (0.002)	253.3
L×A	0.36 (0.005)	157.1
H×C	0.17 (0.989)	150.1
H×A	0.28 (0.438)	157.0
MV game	0.11	
L×C	0.43 (0.000)	45.4
L×A	0.20 (0.000)	58.5
H×C	0.35 (0.000)	48.4
H×A	0.29 (0.000)	58.9

Notes: The entries 0.27 for CH and 0.11 for MV are the fractions of feasible payoff space occupied by CE payoffs; other entries in that column are fractions of observed pair payoffs in the CE zone. (P-values are shown in parentheses for proportional tests with errors clustered at the session level, for the null hypothesis that the observed fraction is no greater than 0.27[0.11] in CH[MV] games.) Last column reports mean Euclidean distance of observed payoff vector from target CE payoff vector.

however, we strongly reject that null hypothesis in both low information treatments. In particular, consistent with H2(a,b), the observed fraction 0.57 of CE-consistent play in the CH×L×C treatment is far larger than the 0.27 random benchmark. Of course, from Figure 5 we know that much of that CE play arises from the two pure NE.

For all MV game treatments, we strongly reject that the observed fraction is no greater that the random proportion 0.11; all p-values are less than 0.001. Again, the largest fraction of payoff pairs in the CE region (0.43) is for counterfactual regret and low information. On the other hand, under Average regret the fraction of pairs in the CE region is larger in the high information (0.29) than in the low information environment (0.20).

Result 2b. *In the CH game, the distance to target CE is driven mainly by collusion in high information treatments, and mainly by pure NE play in the L×C treatment. In the MV game, the distance to target CE is closer under counterfactual regret, with no significant differences across high and low information environments.*

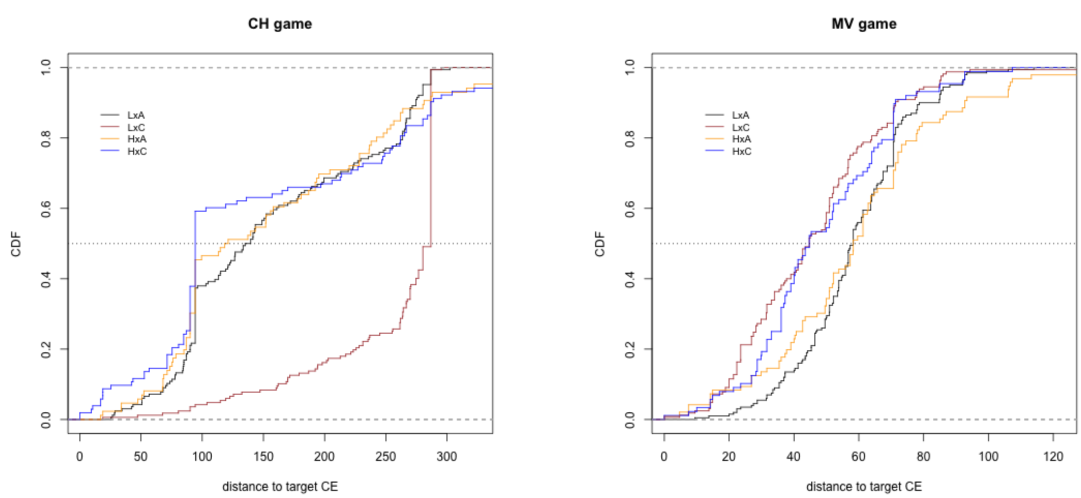


Figure 7: CDF of the distance to target CE (pair data)

This result is foreshadowed by the mean distances reported in last column of Table 7, but it relies mainly on the finer grained analysis displayed in Figure 7. We calculate the Euclidean distance from the target CE payoff for each player pair, and build the cumulative distribution functions (CDFs) shown in the Figure. Its left panel shows that in the CH game, a large fraction of pairs in the L×C treatment have distance 287, which is the distance from either pure NE payoff vector to that of the target CE. In high information treatments, play is closer to the target CE mainly due to a the mass of play near the collusion profile, which has distance 95 from the target CE.

In the MV game (right panel of Figure 7), we observe that play under average regret almost first order stochastically dominates play under counterfactual regret. We test the distance to target CE across regret treatments using a linear regression and clustered standard errors at the session level, and find that the distance to target CE is closer by 11 (p-value of 0.018) under counterfactual regret in the high information treatment, and by 13 (p-value of 0.001) in the low information treatment; see Table 12 in Appendix B.

5.3 Individual adjustment behavior

Result 3. *Individual players in all treatments except CH×H×C (in which collusive behavior abounds) are indeed more likely to switch to an alternative action the larger its positive regret. Also consistent with Hypothesis 3, in most MV game treatments, subjects are less responsive to negative regret.*

Table 8: Maximum Likelihood Estimation

Treatment	Obs	logit	inertia logit		inertia truncated logit		
		β	Δt	β	Δt	β_1	β_2
CH game							
L×C	16,404	2.46	0.48	2.90	0.55	1.47	1.60
		(0.000)	(0.000)	(0.000)	(0.000)	(0.118)	(0.149)
L×A	16,178	0.61	0.48	0.51	0.49	0.39	0.15
		(0.000)	(0.000)	(0.000)	(0.000)	(0.105)	(0.607)
H×C	10,024	-1.44	0.33	-1.42	0.48	-2.06	2.41
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
H×A	8,428	0.50	0.60	0.48	0.76	-0.19	0.81
		(0.000)	(0.000)	(0.000)	(0.000)	(0.087)	(0.000)
MV game							
L×C	16,161	1.64	0.98	1.71	1.00	1.89	-0.40
		(0.000)	(0.000)	(0.000)	(0.076)	(0.000)	(0.001)
L×A	19,496	0.85	0.97	0.88	0.99	0.97	-0.18
		(0.000)	(0.000)	(0.000)	(0.602)	(0.000)	(0.007)
H×C	8,564	1.09	0.97	1.13	1.00	1.34	-0.51
		(0.000)	(0.000)	(0.000)	(0.441)	(0.000)	(0.001)
H×A	9,402	0.91	0.99	0.92	0.98	0.91	0.02
		(0.000)	(0.068)	(0.000)	(0.202)	(0.000)	(0.839)

Notes: Logit specification refers to equation (8); inertia logit to equation (9); and inertia truncated logit uses a multiplicative dummy when the regret is negative. In parentheses are p-values given error clustering at the subject level for the null hypotheses that βs are zero and that $\Delta t = 1$.

This result is obtained from the logit regressions reported in Table 8. Under *logit* specification, we estimate equation (8) for all treatments, and find that the value of β

is strongly positive (p-value < 0.001) in all cases, except the CH game under treatment H$\times$C due to collusive behavior. In specification *inertia logit*, we estimate equation (9), which includes an additional parameter Δt; inertia is present to the degree that its coefficient is less than 1.0 (but positive.) We see very little inertia in the MV game treatments, but see a substantial degree of inertia in CH games. Nevertheless, even in CH games, the β estimates (and significance levels) are not greatly altered by the presence of the additional parameter.

For the third specification, inertia truncated logit, we replace the expression $\beta r_t^i(j,k)$ in equation (9) by $\beta_1 r_t^i(j,k) + \beta_2 r_t^i(j,k) D[r_t^i(j,k) < 0]$, where the dummy variable D is 1 when regret is negative and otherwise is 0. Thus the response to negative regret is captured by the coefficient sum $\beta_1 + \beta_2$. We do not find evidence of full truncation ($\beta_1 + \beta_2 = 0$) in any treatment. However, in 3 of the 4 MV game treatments we do see significant partial offsets as hypothesized: except for an insignificant estimate in the H$\times$A treatment, the coefficient estimates for β_2 are significantly negative.

We also estimate μ in HM response following equation (7). The results support positive response to the regret except in CH $\times$ A treatments (due to the high collusion rate). See Table 13 in Appendix B for details.

5.4 Simulations

To better understand the implications of individual adjustment behavior, we use estimates of logit response models in Table 8 and HM models in Table 13 to perform simulations. We focus here on estimates from low information environments since these environments (a) seem more in the spirit of the relevant adaptive dynamics models, and (b) produce observed outcomes better aligned with CE.

To initialize regret, we begin each simulation with 100 iid action draws. Then we simulate 400 periods with player pairs each adapting their actions as specified in equation (9) using the $(\beta, \Delta t)$ coefficient values reported in the middle columns of Table 8, and the μ in the regression Table 13. Tables 9 and 10 report the joint action distribution from 500 such simulations in the left panel, and the corresponding joint distribution of our human subject sessions on the right. Results are qualitatively similar for the simple logit, the inertia truncated logit models, and the HM models; see Tables 14 and 15 in Appendix C.

Result 4. *Simulations capture salient behavioral differences between counterfactual*

Table 9: Simulations of the CH games: frequency profiles

	Sims: Inertia logit response					**Lab data: Low Information**			
	Average		*Counterfactual*			*Average*		*Counterfactual*	
	L	R	L	R		L	R	L	R
U	0.07	0.17	0.02	0.46	U	0.12	0.22	0.07	0.40
D	0.18	0.58	0.50	0.02	D	0.27	0.39	0.45	0.07

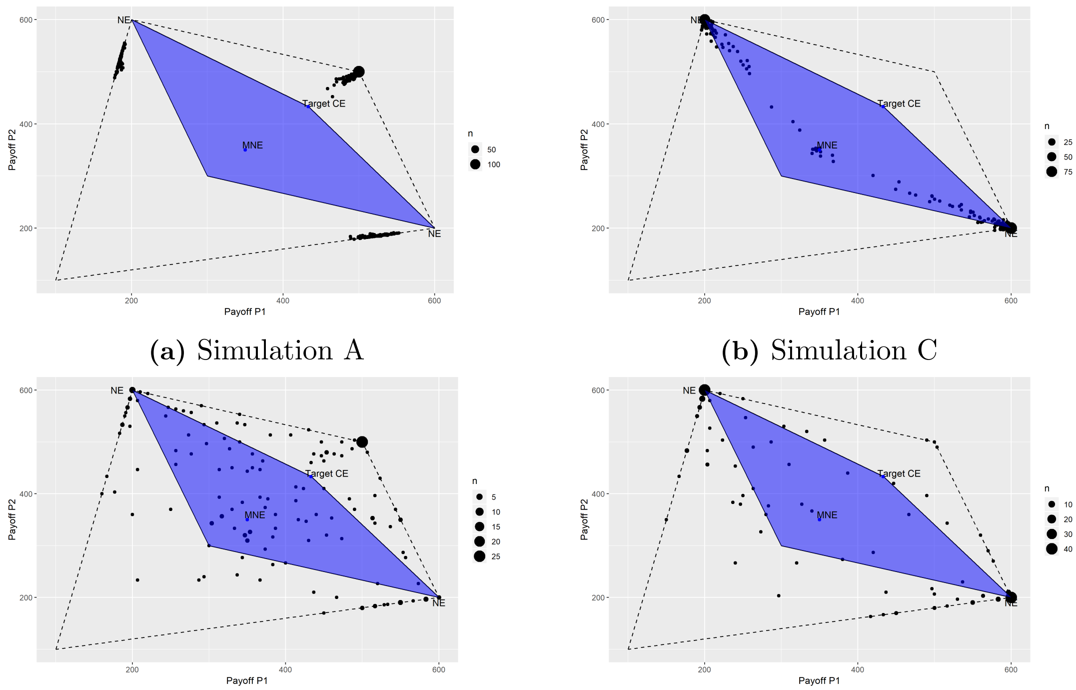
and average regret treatments.

Overall, we find that simulations generally track the joint distribution observed in our experiments across the different forms of regret. In the CH game, the joint distribution is closer to the target CE under average regret, and is closer to the pure NE under counterfactual regret. We also observe that adaptive dynamics under average regret lead to more collusive behavior. The frequency in the cell (D,R) is higher in the simulation (0.58) than in the lab data (0.39). Similarly, our simulations perform well in replicating the action profiles observed in the MV game, where the main diagonal has the lowest frequency of play.

Table 10: Simulations of the MV games: frequency profiles

	Sims: Inertia Logit response							**Lab data: Low Information**					
	Average			*Counterfactual*				*Average*			*Counterfactual*		
	L	C	R	L	C	R		L	C	R	L	C	R
T	0.00	0.17	0.19	0.01	0.17	0.16	T	0.07	0.14	0.13	0.04	0.14	0.12
M	0.15	0.00	0.17	0.16	0.01	0.15	M	0.13	0.07	0.13	0.13	0.05	0.12
D	0.17	0.14	0.00	0.16	0.17	0.01	D	0.12	0.15	0.07	0.18	0.17	0.06

We also present the time average payoff at the pair level for each simulation using the inertia logit response model. Figure 8 shows the time average payoff for each simulation run in the CH game. The simulations converge to either of the two pure NE under counterfactual regret (panel (b)), and move to collusion under average regret (panel(a)). This difference across regret feedback is consistent with what we observe in the experimental data. For the MV game in Figure 9, we find that the time average payoff is closer to the CE under counterfactual regret than average regret, which is also consistent with the data. To obtain a quantitative insight, Tables 16 and 17 in Appendix C follow Table 7 and present the fraction of simulations in the CE and the distance to target for both games, respectively.



(a) Simulation A

(b) Simulation C

(c) Lab L × A

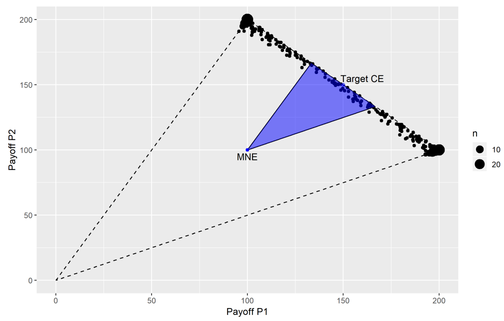
(d) Lab L × C

Figure 8: Simulation runs under inertia logit response model displayed in the CH payoff space.

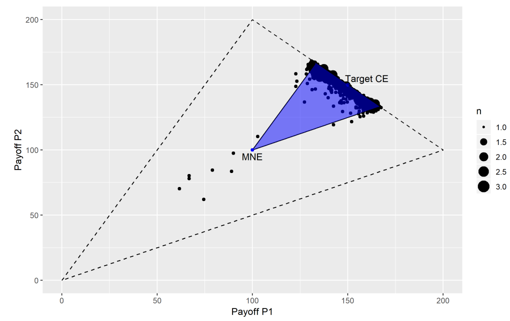
6 Conclusion

Our experiment studies the relevance of correlated equilibria (CE) in two very different bimatrix games. The first, the well known game of Chicken, features two pure Nash equilibria as well as an efficient symmetric "collusive" outcome. The second game, proposed by Moulin and Vial (1978), has no pure strategy equilibrium and offers little scope for collusive behavior. Both games have efficient and fair ("target") CEs that provide a larger total payoff than any NE, and split it evenly. Our experiment provides human players the information they need to easily compute different sorts of regret (Counterfactual or Average) but relatively little other information. Thus we investigate empirically the theoretical possibility, implicit in Foster and Vohra (1998) and Hart and Mas-Colell (2000), that shared experience might enable players to achieve CE even in the absence of exogenous coordinating signals.

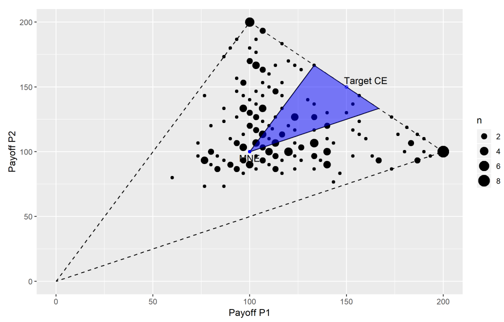
The results are intriguing. We find that, in most treatments, players indeed do respond to regret in a manner consistent with theory. In the Chicken game, the modal outcomes are either pure NE or collusion, depending on the information and regret



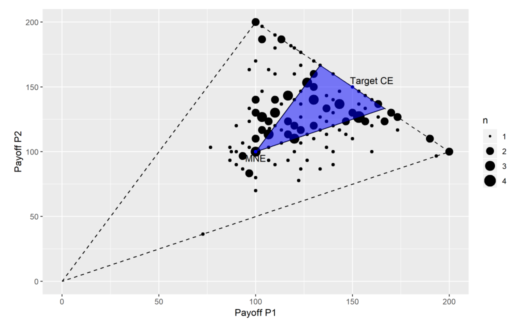
(a) Simulation A



(b) Simulation C



(c) Lab L×A



(d) Lab L×C

Figure 9: Simulation runs under inertia logit response model displayed in the MV payoff space.

treatments. In the MV game we indeed see behavior that more closely resembles the target CE than the unique (mixed) NE.

We do not regard our experiment as definitive, but rather as opening new ways to investigate strategic interactions. Our Low (and High, i.e., moderate) information user interfaces and other novel design features can be adapted to study other games with interesting CE, including trimatrix games with some players having two and others having three possible pure actions. Our payoff space analysis can help keep manageable the number of dimensions, and our simulation techniques can help pre-explore the vast set of possible games and treatments. We hope future researchers will find these innovation useful in investigating the ability of CE to explain outcomes in the absence of exogenous signals, in other sorts of strategic interaction, and in studying the convergence of uncoupled learning models.

7 Acknowledgements

We gratefully acknowledge funding for this project from Early Career Grants, Monash Business School. For comments and feedback we are indebted to Aleksandr Alekseev, Tim Cason, John Duffy, Nick Feltovich, Misha Freer, Evan Friedman, Sergiu Hart, Heinrich Nax, Jan Potters, Vjollca Sadiraj, Tridib Sharma, Rakesh Vohra, and participants in numerous seminars and workshops. This paper is dedicated to the memory of Bill Sandholm, whose encouragement and suggestions in 2017 eventually helped us develop the experiment reported herein.

8 Declarations of Interest

None. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Anbarcı, Nejat, Nick Feltovich, and Mehmet Y Gürdal, "Payoff inequity reduces the effectiveness of correlated-equilibrium recommendations," *European Economic Review*, 2018, *108*, 172–190.

Arifovic, Jasmina, Joshua F Boitnott, and John Duffy, "Learning correlated equilibria: An evolutionary approach," *Journal of Economic Behavior & Organization*, 2019, *157*, 171–190.

Aumann, Robert J, "Subjectivity and correlation in randomized strategies," *Journal of mathematical Economics*, 1974, *1* (1), 67–96.

_ , "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica: Journal of the Econometric Society*, 1987, pp. 1–18.

Bone, John, Michalis Droiuvelis, and Indrajit Ray, "Coordination in 2 x 2 games by following recommendations from correlated equilibria," 2013.

Brandenburger, Adam and Eddie Dekel, "Rationalizability and correlated equilibria," *Econometrica: Journal of the Econometric Society*, 1987, pp. 1391–1402.

Cason, Timothy N and Tridib Sharma, "Recommended play and correlated equilibria: an experimental study," *Economic Theory*, 2007, *33* (1), 11–27.

_ , _ , and Radovan Vadovič, "Correlated beliefs: Predicting outcomes in 2× 2 games," *Games and Economic Behavior*, 2020, *122*, 256–276.

Duffy, John and Nick Feltovich, "Correlated equilibria, good and bad: an experimental study," *International Economic Review*, 2010, *51* (3), 701–721.

_ , Ernest K Lai, and Wooyoung Lim, "Coordination via correlation: An experimental study," *Economic Theory*, 2017, *64* (2), 265–304.

Forges, Francoise and James Peck, "Correlated equilibrium and sunspot equilibrium," *Economic Theory*, 1995, *5* (1), 33–50.

Foster, Dean and Hobart Peyton Young, "Regret testing: Learning to play Nash equilibrium without knowing you have an opponent," *Theoretical Economics*, 2006, *1* (3), 341–367.

Foster, Dean P and Rakesh V Vohra, "Calibrated learning and correlated equilibrium," *Games and Economic Behavior*, 1997, *21* (1-2), 40.

_ and _ , "Asymptotic Calibration," *Biometrika*, 1998, *85*, 379–390.

Fudenberg, Drew and David K Levine, "Conditional universal consistency," *Games and Economic Behavior*, 1999, *29* (1-2), 104–130.

Georgalos, Konstantinos, Indrajit Ray, and Sonali SenGupta, "Nash versus coarse correlation," *Experimental Economics*, 2020, pp. 1–27.

Hart, Sergiu and Andreu Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, 2000, *68* (5), 1127–1150.

_ and _ , "A reinforcement procedure leading to correlated equilibrium," in "Economics essays," Springer, 2001, pp. 181–200.

Metzger, Lars P, "Evolution and correlated equilibrium," *Journal of Evolutionary Economics*, 2018, *28* (2), 333–346.

Moreno, Diego and John Wooders, "An experimental study of communication and coordination in noncooperative games," *Games and Economic Behavior*, 1998, *24* (1-2), 47–76.

Moulin, Hervé and J-P Vial, "Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon," *International Journal of Game Theory*, 1978, *7* (3-4), 201–221.

Nax, Heinrich H, Maxwell N Burton-Chellew, Stuart A West, and H Peyton Young, "Learning in a black box," *Journal of Economic Behavior & Organization*, 2016, *127*, 1–15.

Palfrey, Thomas R and Kirill Pogorelskiy, "Communication among voters benefits the majority party," *The Economic Journal*, 2019, *129* (618), 961–990.

Pradelski, Bary SR and H Peyton Young, "Learning efficient Nash equilibria in distributed systems," *Games and Economic behavior*, 2012, *75* (2), 882–897.

Young, H Peyton, "Learning by trial and error," *Games and economic behavior*, 2009, *65* (2), 626–643.

9 Appendix

A: Calculate Corr Equilibrium

	L	C	R
T	0	1	2
M	2	0	1
B	1	2	0

Table 11: Moulin-Vial game

For row player

$$q_{11}(2-0)+q_{12}(0-1)+q_{13}(1-2) \leq 0 \quad (10)$$
$$q_{11}(1-0)+q_{12}(2-1)+q_{13}(0-2) \leq 0 \quad (11)$$
$$q_{21}(0-2)+q_{22}(1-0)+q_{23}(2-1) \leq 0 \quad (12)$$
$$q_{21}(1-2)+q_{22}(2-0)+q_{23}(0-1) \leq 0 \quad (13)$$
$$q_{31}(0-1)+q_{32}(1-2)+q_{33}(2-0) \leq 0 \quad (14)$$
$$q_{31}(2-1)+q_{32}(0-2)+q_{33}(1-0) \leq 0 \quad (15)$$

For column player

$$q_{11}(2-0)+q_{21}(0-1)+q_{31}(1-2) \leq 0 \quad (16)$$
$$q_{11}(1-0)+q_{21}(2-1)+q_{31}(0-2) \leq 0 \quad (17)$$
$$q_{12}(0-2)+q_{22}(1-0)+q_{32}(2-1) \leq 0 \quad (18)$$
$$q_{12}(1-2)+q_{22}(2-0)+q_{32}(0-1) \leq 0 \quad (19)$$
$$q_{13}(0-1)+q_{23}(1-2)+q_{33}(2-0) \leq 0 \quad (20)$$
$$q_{13}(2-1)+q_{23}(0-2)+q_{33}(1-0) \leq 0 \quad (21)$$

And, $\sum q_s = 1$

B: Additional Regressions

Table 12: Distance to target CE in MV game (OLS)

	(I) Low Information	(II) High Information
Intercept	58.53	58.88
	(.000)	(.000)
Counterfactual	−13.13	−10.52
	(.001)	(.018)
R^2	0.10	0.05
N	365	184

Notes: p-values in parenthesis. Standard errors clustered at the session level. *Counterfactual* is a dummy variable that takes the value of one when the regret is counterfactual, and zero when regret is historical average.

Table 13: Estimating μ in HM model with data.

	(I) CH games	(II) MV games
$\frac{1}{\mu}$	0.0003***	0.0044***
	(0.000)	(0.000)
A_regret	-0.0004***	-0.0003
	(0.000)	(0.000)
H_info	-0.0001	0.0002
	(0.000)	(0.000)
*A*H*	0.0001	0.0000
R^2	0.003	0.339
N	51,034	53,634

Notes: Linear regressions that estimate μ following equation (7). Standard errors are listed in parenthesis and are clustered at the session level. Independent variables are A_regret which takes the value of 1 if the regret information is average and zero otherwise; H_info which takes the value of one if the treatment is high information, and zero otherwise. The dependent variable is the probability of switching from the current choice to another alternative. The regret term independent variable refers to the regret of the alternative in CH games, and the best regret of the two alternatives in MV games. The estimation in CH games reflects the inertia, while in MV games the actual inertia is between the theoretical minimum $\mu = 400$ and 2×the estimated $\frac{1}{\mu}$.

C: Simulation Details

Tables 14 and 15 show the joint distribution for all four dynamic models. The results with HM response and average regret converges to a uniform mixed NE, which is different from the other three logit models. One possible reason is the negative μ caused by the high collusion rate. However, the HM simulation is not far from the experimental result, just from an opposite direction.

Table 14: Simulation joint distribution CH games

Logit Response				
	Average		*Counterfactual*	
	L	R	L	R
U	0.09	0.16	0.11	0.38
D	0.18	0.56	0.40	0.11

Inertia Logit				
	Average		*Counterfactual*	
	L	R	L	R
U	0.07	0.17	0.02	0.46
D	0.18	0.58	0.50	0.02

Inertia Truncated Logit				
	Average		*Counterfactual*	
	L	R	L	R
U	0.06	0.19	0.02	0.48
D	0.15	0.61	0.48	0.02

HM Response				
	Average		*Counterfactual*	
	L	R	L	R
U	0.24	0.26	0.04	0.47
D	0.26	0.24	0.44	0.05

Following the analysis summarized in Table 7, we also present the fraction of time average payoffs in the CE region as well as the distance to the target CE for CH games. Table 16 presents the logit models estimated above, as well as the linear HM algorithm, equation (7). The simulations show that under average regret, the time average payoff cannot be considered a CE because of the high frequency in the cell (D,R). The only except is from the HM model, where pairs converge to mixed NE inside the CE region. For counterfactual regret, the fraction of play in the CE region varies from 0.52 to 0.59 across all models, which follows human behavior presented in Table 7. The distance to target CE is also quite similar to the data, while the slight difference is mainly caused by the better-convergence in the simulation data.

Simulations for the MV game are reported in Table 17. We find that counterfactual regret yields a higher fraction of play in the CE region than average regret, and the distance to target CE is also quite different. The simulation shows a clear distinction between the two regrets, while the data do not. The main reason is still the better convergence in the simulations. With a much lower error rate in the simulations, the distinction between the two regrets can finally be revealed. The HM response likewise similar to the other specifications, though the fraction of play in the CE region (0.67)

Table 15: Simulation joint distribution MV games

Logit Response

	Average			*Counterfactual*		
	L	C	R	L	C	R
T	0.00	0.18	0.19	0.01	0.16	0.15
M	0.13	0.00	0.13	0.17	0.01	0.15
D	0.19	0.17	0.00	0.17	0.16	0.01

Inertia Logit

	Average			*Counterfactual*		
	L	C	R	L	C	R
T	0.00	0.17	0.19	0.01	0.17	0.16
M	0.15	0.00	0.17	0.16	0.01	0.15
D	0.17	0.14	0.00	0.16	0.17	0.01

Inertia Truncated Logit

	Average			*Counterfactual*		
	L	C	R	L	C	R
T	0.00	0.19	0.16	0.02	0.16	0.16
M	0.15	0.00	0.17	0.16	0.02	0.15
D	0.17	0.16	0.00	0.16	0.16	0.02

HM Response

	Average			*Counterfactual*		
	L	C	R	L	C	R
T	0.01	0.16	0.17	0.00	0.15	0.15
M	0.16	0.01	0.18	0.16	0.00	0.19
D	0.14	0.15	0.01	0.16	0.17	0.00

Table 16: Simulation analysis of CH game

Spec	Regret	Fraction in CE	Distance to Target CE
HM Res	C	0.58	267.6
	A	0.52	283.8
Logit	C	0.52	264.2
	A	0	165.9
I Logit	C	0.59	268.7
	A	0	166.1
IT Logit	C	0.52	275.6
	A	0	158.2
Lab Data	L×C	0.57	253.3
	L×A	0.36	157.1

Notes: Distance is calculated by the Euclidean distance between target NE and the average payoff vector at the pair level.

is smaller than in other models.

Table 17: Simulation analysis of MV game

Spec	Regret	Fraction in CE	Distance to Target CE
HM Res	C	0.67	18.1
	A	0.11	57.8
Logit	C	0.86	17.4
	A	0.08	59.8
I Logit	C	0.88	15.3
	A	0.09	59.2
IT Logit	C	0.87	17.9
	A	0.08	59.0
Lab Data	L×C	0.43	45.4
	L×A	0.20	58.5

Notes: Distance is calculated by the Euclidean distance between target NE and the average payoff vector at the pair level.

D: Simulation under High Information Parameters

This section shows the simulation results using the parameters estimated in the high information treatment. In CH games, the inertia truncated logit model returns a joint density closer to uniform random mixtures rather than correlated equilibrium, possibly due to the large β_2 related to negative regret. The other models, however, show similar pattern compared to what we have found in the low information treatments.

Table 18: Simulation joint distribution CH games with H information

Logit Response

	Average L	*Average* R	*Counterfactual* L	*Counterfactual* R
U	0.07	0.21	0.34	0.13
D	0.19	0.53	0.13	0.40

Inertia Logit

	Average L	*Average* R	*Counterfactual* L	*Counterfactual* R
U	0.06	0.17	0.49	0.02
D	0.18	0.59	0.02	0.47

Inertia Truncated Logit

	Average L	*Average* R	*Counterfactual* L	*Counterfactual* R
U	0.31	0.20	0.38	0.20
D	0.21	0.28	0.19	0.23

HM Response

	Average L	*Average* R	*Counterfactual* L	*Counterfactual* R
U	0.23	0.24	0.06	0.42
D	0.26	0.26	0.45	0.07

Table 19: Simulation analysis of CH game with H information

Spec	Regret	Fraction in CE	Distance to Target CE
HM Res	C	0.57	262.6
	A	0.51	278.8
Logit	C	0.26	210.3
	A	0	174.4
I Logit	C	0.03	270.9
	A	0	161.5
IT Logit	C	0.25	308.1
	A	0.28	281.2

Table 20: Simulation joint distribution MV games with H information

Logit Response

	Average L	*Average* C	*Average* R	*Counterfactual* L	*Counterfactual* C	*Counterfactual* R
T	0.00	0.15	0.16	0.01	0.17	0.16
M	0.17	0.00	0.17	0.15	0.01	0.14
D	0.16	0.18	0.00	0.17	0.18	0.01

Inertia Logit

	Average L	*Average* C	*Average* R	*Counterfactual* L	*Counterfactual* C	*Counterfactual* R
T	0.00	0.18	0.18	0.01	0.17	0.15
M	0.17	0.00	0.16	0.15	0.01	0.16
D	0.15	0.15	0.00	0.17	0.18	0.01

Inertia Truncated Logit

	Average L	*Average* C	*Average* R	*Counterfactual* L	*Counterfactual* C	*Counterfactual* R
T	0.00	0.16	0.16	0.01	0.17	0.16
M	0.17	0.00	0.16	0.16	0.01	0.15
D	0.19	0.17	0.00	0.16	0.15	0.01

HM Response

	Average L	*Average* C	*Average* R	*Counterfactual* L	*Counterfactual* C	*Counterfactual* R
T	0.01	0.16	0.18	0.00	0.17	0.18
M	0.15	0.01	0.14	0.16	0.00	0.17
D	0.15	0.19	0.01	0.15	0.16	0.00

Table 21: Simulation analysis of MV game with H information

Spec	Regret	Fraction in CE	Distance to Target CE
HM Res	C	0.67	17.5
	A	0.06	59.4
Logit	C	0.89	17.7
	A	0.09	58.6
I Logit	C	0.91	14.1
	A	0.07	58.7
IT Logit	C	0.87	18.5
	A	0.09	59.1