Games and Economic Behavior 130 (2021) 352–368

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

The determinants of efficient behavior in coordination games

Pedro Dal Bó [a,*], Guillaume R. Fréchette [b], Jeongbin Kim [c,1]

[a] Brown University & NBER, United States of America
[b] New York University, United States of America
[c] National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 21 May 2020
Available online 6 September 2021

JEL classification:
C9
D7

Keywords:
Coordination games
Strategic uncertainty
Equilibrium selection

ABSTRACT

We study the determinants of efficient behavior in stag hunt games (2×2 symmetric coordination games with Pareto ranked equilibria) using data from eight previous experiments on stag hunt games and a new experiment that allows for a more systematic variation of parameters. We find that subjects do not necessarily play the efficient action (stag), stressing the importance of strategic uncertainty in coordination games. While the frequency of playing stag is greater when stag is risk dominant, there is still large variation in behavior that cannot be explained by risk dominance. Part of this variation is explained by the risk arising from strategic uncertainty that we measure with the size of the basin of attraction of stag. We also explore the importance of other determinants of efficient behavior and we show that the results are robust to paying subjects using the lottery method in an attempt to induce risk neutral preferences.

1. Introduction

The study of coordination games has a long history as many situations of interest present a coordination component, for example: the choice of technologies that require a threshold number of users to be sustainable, currency attacks, bank runs, asset price bubbles, cooperation in repeated games, etc. In such examples, agents may face strategic uncertainty; that is, they may be uncertain about how the other agents will respond to the multiplicity of equilibria, even when they have complete information about the environment.

A simple coordination game that captures the main forces present in the previous examples is the well-known stag hunt game: a two-player and two-choice game with Pareto ranked equilibria. That game features two Nash equilibria in pure strategies in which either both players select stag, or both players select hare; with stag being socially optimal (payoff dominant). Such a simple game allows the study of conditions that lead people to coordinate on the efficient equilibrium.

The first experimental study of the stag hunt game, Cooper et al. (1992), focuses on a stag hunt game in which hare is risk dominant (that is, hare is the best response to the belief that the other player is randomizing 50-50 between stag and hare). They find that, absent communication, an overwhelming fraction of choices are in line with the risk dominant choice of hare. This is consistent with the idea that people may coordinate on the action most robust to strategic uncertainty.

* Corresponding author.
E-mail address: pdalbo@brown.edu (P. Dal Bó).

1 We thank all the authors who made available the data that we use from previous articles, Hui Wen Ng for her assistance running experiments and Anna Aizer for useful comments. We also thank the editors and referees for their helpful comments and suggestions. Fréchette gratefully acknowledges financial support from the Center for Experimental Social Science (CESS), the C.V. Starr Center, and the National Science Foundation via grant SES-1558857.

https://doi.org/10.1016/j.geb.2021.08.010
0899-8256/

Table 1
Stag hunt game - row player's payoffs.

	Original			Normalized	
	hare	stag		hare	stag
hare	P	T	hare	$\frac{P-P}{R-P}=0$	$\frac{T-P}{R-P}=1-\Lambda$
stag	S	R	stag	$\frac{S-P}{R-P}=-\lambda$	$\frac{R-P}{R-P}=1$

Relatedly, experiments on the minimum effort game, starting with Van Huyck et al. (1990), find a strong tendency for behavior to quickly settle on the minimum effort, where strategic risk is minimized (as opposed to the maximal effort—payoff dominant—equilibrium). Despite other studies that followed with mixed results, see for example Straub (1995), Battalio et al. (2001) and Schmidt et al. (2003), these early results created a strong notion that risk dominance was the key determinant of behavior in such coordination games.

In this paper, we return to stag hunt games for a systematic assessment of the determinants of efficient behavior (playing stag) using two data sets. First, we study behavior in the metadata from eight previous experiments on stag hunt games. However, as will become clear, those studies show a surprisingly limited amount of treatment variations. Second, we study behavior from a new experiment that allows us to easily explore more parameter combinations than in the previous experiments. In each round of this experiment, subjects participate in sixteen simultaneous stag hunt games with different payoff parameters, allowing us to explore many payoff combinations efficiently. Moreover, in some sessions we use the lottery procedure introduced by Roth and Malouf (1979) in an attempt to affect risk preferences and explore their impact on behavior. Third, we conduct additional experiments aimed at understanding whether playing 16 simultaneous games affects choices.

We find that people do not necessarily coordinate on the efficient equilibrium. In fact, for some treatments, only a very small minority plays stag. The fact that payoff dominance is not used by the subjects as an equilibrium selection criterion suggests that strategic uncertainty may be important in coordination games. As such, one may believe that agents would choose actions corresponding to the equilibrium most robust to strategic uncertainty, that is, the risk dominant action (Harsanyi and Selten (1988)). While we find that stag is more prevalent, on average, when it is risk dominant; it is not always the case that a majority of people coordinate on the risk dominant equilibrium. Risk dominance, on its own, leaves much variation in behavior unexplained. We find that part of this variation can be explained by a continuous measure of the risks arising from strategic uncertainty. This measure is the size of the basin of attraction of stag, which is the maximum probability of the other player choosing hare such that playing stag is still a best response. The greater this number, the more robust playing stag is to strategic uncertainty.[2] We find that the share of subjects choosing stag increases in the size of its basin of attraction. We also show the importance of other determinants of behavior studied in the literature such as the optimization premium (Battalio et al. (2001)) and hare's relative riskiness (Dubois et al. (2012)). Furthermore, these results hold whether we pay subjects for points accumulated or via the lottery method, suggesting that either risk preferences may not be an important driver of behavior in this application or that the lottery method did not affect risk attitudes.

Interestingly, although the effect of the size of the basin of attraction of stag on efficient behavior is found both in the metadata from the previous literature and in the experiments using our new design, the exact relation is different. The prevalence of stag for intermediate sizes of the basin of attraction of stag is lower in the new experiment than in the earlier experiments. This suggests that behavior in a coordination game may depend not only on the parameters of the game, but also on whether subjects participate in several games simultaneously or only one game at a time. We provide evidence that this difference is due to subjects displaying more noise in choices when they participate in multiple games simultaneously.

2. Theoretical background

The stag hunt game is a two-player game with two actions, stag and hare, with the payoffs as shown in Table 1 (Original) with the constraint on payoffs that $T<R>P>S$. Note that $(stag, stag)$ is a Nash equilibrium given that $T<R$, but $(hare, hare)$ is also a Nash equilibrium given that $S<P$. Given that $R>P$, the former equilibrium results in higher payoffs than the latter one. Following Harsanyi and Selten (1988), we say that $(stag, stag)$ is the payoff dominant (or Pareto efficient) equilibrium and stag is the payoff dominant action. There is also a mixed strategy Nash equilibrium in which subjects play hare with probability $\frac{1}{1+\frac{P-S}{R-T}}$.

Any stag hunt game with four parameters R, S, T, P as in the left panel of Table 1 can be normalized by linear transformation of payoffs to a game with only two parameters, Λ and λ, as in the right panel of Table 1 (Normalized). The parameter Λ denotes the loss arising from an unilateral deviation from the efficient equilibrium, while the parameter λ

[2] There is a connection here with the study of cooperation in repeated prisoner's dilemma games. If the infinitely repeated game is suitably simplified, by focusing on a cooperative strategy (grim) and a defecting strategy (always defect), it can be reduced to a stag hunt game. Thus, one can identify parameters for which cooperation can be supported as part of a risk dominant and payoff dominant equilibrium versus others where only defection can be risk dominant, see Blonski and Spagnolo (2015). Dal Bó and Fréchette (2011) and Dal Bó and Fréchette (2018) show that variation in cooperation rates is related to the size of the basin of attraction of the strategies in the simplified game. It has also been found that the basin of attraction is an important determinant of behavior in other games, see Healy (2016), Calford and Oprea (2017), Embrey et al. (2017), Vespa and Wilson (2017), Kartal and Müller (2018), and Castillo and Dianat (2018).

Table 2
Stag hunt games - row player's payoffs.

Example 1	hare	stag	Example 2	hare	stag
hare	0	-1	hare	0	-1
stag	-1	1	stag	-100	1

denotes the loss arising from an unilateral deviation from the inefficient equilibrium. This normalization will allow us to compare behavior across stag hunt experiments while keeping track of only two payoff parameters, Λ and λ, instead of the four original parameters.[3]

How should we expect people to behave in the stag hunt game? Previous authors, see for example Luce and Raiffa (1957), Schelling (1960), and Harsanyi and Selten (1988), have theorized that people would coordinate on the efficient equilibrium, in this case (*stag*, *stag*). This is quite intuitive for a game as the one shown in the left panel of Table 2 (example 1), but may be less so in the game shown in the right panel (example 2). For the game on the right, even a small amount of uncertainty about the action of the other player can make stag a sub-optimal choice. In other words, (*stag*, *stag*) is not very robust to strategic uncertainty in the second example.

The robustness to strategic uncertainty of the equilibrium (*stag*, *stag*) can be measured by the maximum probability of the other subject playing hare that still makes stag a best response. This number is provided by the probability of hare in the mixed strategy Nash equilibrium and is usually referred to as the size of the basin of attraction of stag. Under normalized payoffs, the size of the basin of attraction of stag is equal to $\frac{\Lambda}{\Lambda+\lambda}$. Note that, intuitively, this number is decreasing in λ and increasing in Λ.[4] Following Harsanyi and Selten (1988), we say that stag is risk dominant if its basin of attraction is greater than one half. If that is the case, (*stag*, *stag*) is more robust to strategic uncertainty than (*hare*, *hare*). Harsanyi and Selten (1988) proposed risk dominance as an alternative equilibrium selection criterion. The idea that people may coordinate on the risk dominant equilibrium received support from evolutionary theories (see Kandori et al. (1993) and Young (1993)).

While the previous experimental literature on coordination games has shown that subjects do not necessarily coordinate on the efficient equilibrium (see Cooper et al. (1990), Van Huyck et al. (1990), and Cooper et al. (1992)), the literature has not yet provided a clear answer to the issue of when people would coordinate on the efficient equilibrium. In particular, we seek to answer the following questions. Is it the case that people are more likely to play stag if it is risk dominant? Does the prevalence of stag depend on how robust it is to strategic uncertainty (that is, the size of its basin of attraction)? Are there other determinants of efficient behavior? Moreover, are the answers to these questions different when we use the lottery method in an attempt to induce risk neutrality?

3. Determinants of efficient behavior: previous experiments

We have identified nine published articles with previous stag hunt experiments that are amenable to be analyzed jointly; of which we were able to obtain the data from eight of them.[5] The collected data satisfies the following conditions: (1) 2×2 stag hunt game, (2) no pre-play communication, and (3) using non-fixed matching across periods.[6]

We refer to this data set as the metadata for simplicity, even though it is a collection of raw data sets rather than a collection of aggregated data sets as in a typical metadata.

Table 3 summarizes the treatments in the previous experiments that satisfy the conditions described above. Some of the papers have treatments that do not fit our criteria, e.g., treatments with pre-play communication or with fixed matching throughout the experiment, and those treatments are not included in our analysis. We have data from eight articles, involving 18 different treatments (combinations of the four payoff parameters T, R, P, and S), with 90 experimental sessions and 970 subjects. The vast majority of treatments are such that hare is risk dominant (14 out of 18 treatments) and in only two treatments stag is risk dominant. That is, in most treatments from previous articles there is a tension between payoff dominance and risk dominance.

We study behavior in period 1 as well as in period 8. The latter period is the largest period with observations in every treatment, as the experiment with the smallest number of periods is Schmidt et al. (2003) with 8 periods. Focusing on period 8 allows us to study behavior across treatments after subjects have gained some experience.

[3] Such normalization is common in game theory, but we have not encountered it in experimental papers on the Stag Hunt game. This may in part explain why, as we will see later, there is very little variation in *normalized* parameters across experiments.

[4] An alternative measure of the robustness of (*stag*, *stag*) to strategic uncertainty is $R = ln\frac{\lambda}{\Lambda}$—see Selten (1995) and Schmidt et al. (2003). This measure ranks games as the size of the basin of attraction of stag but the metric is different.

[5] The eight articles for which we have data are Cooper et al. (1992), Straub (1995), Battalio et al. (2001), Clark et al. (2001), Duffy and Feltovich (2002), Schmidt et al. (2003), Dubois et al. (2012), and Feltovich et al. (2012). The data from Charness (2000) is no-longer retrievable.

[6] In particular, Clark et al. (2001), Schmidt et al. (2003), and Straub (1995) use the perfect stranger matching. In Cooper et al. (1992), subjects play against every other player twice: once as a row player and once as a column player. Battalio et al. (2001), Dubois et al. (2012) and Feltovich et al. (2012) use random matching across periods. In Duffy and Feltovich (2002), subjects are assigned to the role of a row or a column player which remain fixed throughout the experiment and play with every other subject of the opposite role.

Table 3
Treatment parameters in prior experiments.

	Λ	λ	Basin of stag	Sessions	Subjects	Periods
Battalio et al. (2001)				**24**	**192**	
	0.091	0.364	0.2	8	64	75
	0.2	0.8	0.2	8	64	75
	2	8	0.2	8	64	75
Clark et al. (2001)				**5**	**100**	
	0.333	2.333	0.125	2	40	10
	1	4	0.2	2	40	10
	3	9	0.25	1	20	10
Cooper et al. (1992)	1	4	0.20	**3**	**30**	22
Dubois et al. (2012)				**24**	**192**	
	0.091	0.364	0.2	8	64	75
	0.375	1.5	0.2	8	64	75
	0.375	1.5	0.2	8	64	75
Duffy and Feltovich (2002)	1	3	0.25	**3**	**60**	10
Feltovich et al. (2012)				**10**	**186**	
	1	2.2	0.313	6	90	20
	2	1	0.667	4	96	40
Schmidt et al. (2003)				**16**	**160**	
	0.5	1.5	0.25	4	40	8
	1	1	0.5	4	40	8
	1	1	0.5	4	40	8
	1	3	0.25	4	40	8
Straub (1995)				**5**	**50**	
	0.2	0.4	0.333	1	10	9
	1	0.5	0.667	1	10	9
	1	1	0.5	1	10	9
	1	3	0.25	1	10	9
	1	4	0.2	1	10	9
Total				**90**	**970**	

First, we consider the effect of payoff and risk dominance. Second, we consider the role of strategic uncertainty as measured by the basin of attraction of stag. Third, we study other determinants of behavior such as the optimization premium (Battalio et al. (2001)) and hare's relative riskiness (Dubois et al. (2012)), and impacts of experimental design features (such as the use of the lottery method for payments). However, we note that the treatment of each of these questions in the meta analysis is brief, because the first finding is that despite many varied stage games having been studied, there is surprisingly little systematic variation in some of the studied determinants of behavior. In particular, while the basin of attraction of stag goes from $\frac{1}{8}$ to $\frac{2}{3}$ across treatments, there is actually very limited variation in this dimension as 72% of treatments have a size of the basin of stag in a small interval (between $\frac{1}{5}$ to $\frac{1}{3}$).

Payoff dominance: Across all papers and all treatments, 67% of subjects chose stag in period 1 while 55% do so in period 8. Fig. 1 shows the distribution of the prevalence of stag at the session level (separating treatments depending on whether they are risk dominant or not). We see that, by period 8, there are sessions going from 0% of stag to 100%, showing that people do not necessarily coordinate on the payoff dominant equilibrium.

It is interesting to note that there are differences across articles in the observed prevalence of stag which led them to reach different conclusions with respect to the importance of payoff dominance. For instance, the treatments in Cooper et al. (1992) resulted in 23% of stag in period 8 and only 3% in their last period, which leads them to conclude that "coordination failures always occur." Instead, the treatments in Schmidt et al. (2003) resulted in 66% of stag in period 8 (the last period in that experiment). This led to their conclusion that "our results could be seen as supporting Harsanyi and Selten's original assertion that players should trust one another to play the payoff dominant equilibrium."

Risk dominance: As Fig. 1 shows, subjects are significantly more likely to choose stag when it is risk dominant. In period 1, there is 64% of stag when it is not risk dominant versus 91% when it is risk dominant (p-value < 0.01).[7] In period 8, there is 50% of stag when it is not risk dominant versus 96% when it is risk dominant (p-value < 0.01). However, there is great variation in behavior for treatments in which stag is not risk dominant: the prevalence of stag varies between 0% and 100% across sessions in period 8 (see Fig. 1).

[7] We compute significance from a probit analysis of choices with clustering at the article level. This is meant to account for potential correlations between observations of a given article that could result from uncontrolled elements such as the specific instructions, the interface, or other.

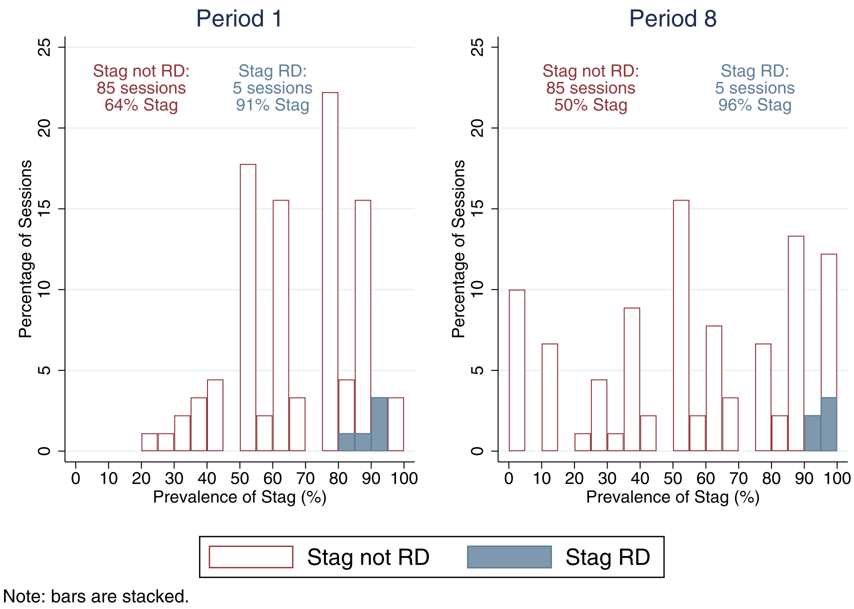


Fig. 1. Meta-analysis: distribution of the prevalence of stag by session.

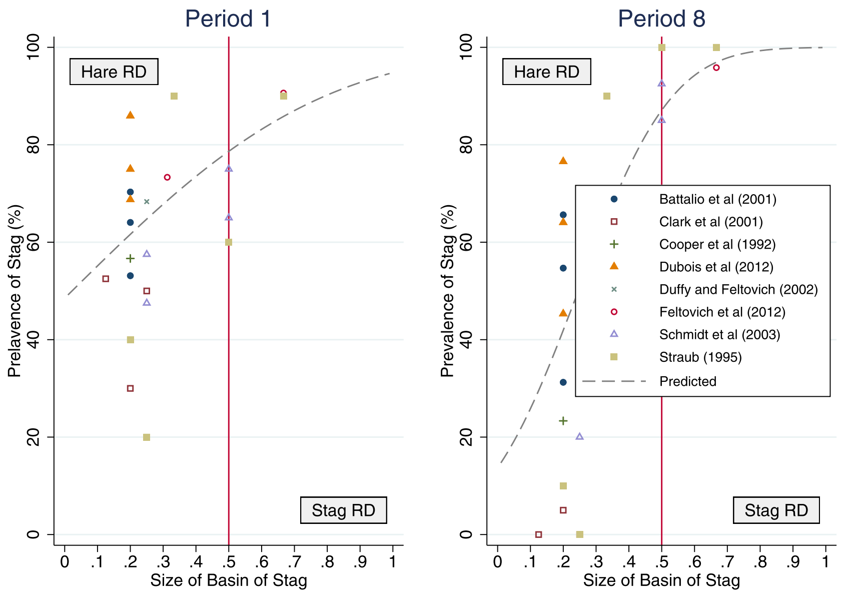


Fig. 2. Meta-analysis: Relation between the prevalence of stag and the size of the basin.

Strategic uncertainty (beyond risk dominance): We study whether the observed variation in the prevalence of stag across treatments can be accounted for by the robustness of the efficient equilibrium $(stag, stag)$ to strategic uncertainty in addition to risk dominance. As discussed in Section 2, we measure robustness to strategic uncertainty by the size of the basin of attraction of stag, which is equal to $\frac{\Lambda}{\Lambda+\lambda}$.

Fig. 2 shows the prevalence of stag in each treatment and article in the metadata for periods 1 and 8 as a function of the size of the basin of attraction of stag. The dashed line is the predicted prevalence resulting from estimating a simple probit of stag on the size of the basin.

Overall, the prevalence of stag is positively correlated with its basin of attraction: as the basin of attraction increases, the prevalence of stag also increases. This relation is present from the onset, but becomes more pronounced with experience (see columns 1 and 2 in Table 4).

Table 4
Meta-analysis. Determinants of stag (probit analysis—marginal effects).

	Period 1	Period 8	Period 1	Period 8	Period 1	Period 8
Stag RD (d)	0.19***	-0.08	0.06	-0.45***	0.09	-0.42*
	(0.043)	(0.162)	(0.099)	(0.146)	(0.092)	(0.224)
Basin of Stag	0.28	1.82***	0.51**	2.47***	0.62***	3.08***
	(0.246)	(0.602)	(0.229)	(0.545)	(0.155)	(0.787)
OP			-0.01**	-0.03	-0.02***	-0.04***
			(0.005)	(0.019)	(0.004)	(0.010)
Hare's RR			0.26	0.56***	0.09	0.06
			(0.158)	(0.120)	(0.161)	(0.228)
Lottery (d)			0.02	0.10	0.05	0.20*
			(0.072)	(0.110)	(0.067)	(0.105)
# Interactions					0.01***	0.04***
					(0.003)	(0.013)
Observations	970	970	970	970	970	970

(d) denotes dummy variable, effect of change from 0 to 1 is reported.
OP denotes Optimization Premium ($\Lambda + \lambda$). Hare's RR denotes hare's relative riskiness $\left(\frac{|1-\Lambda|}{1+\lambda}\right)$.
Lottery denotes Roth-Malouf lottery was used.
Interactions denotes the expected number of interactions with the same subject.
Standard errors clustered at the article level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Fig. 2 also suggests that the size of the basin of attraction on its own cannot account for all of the observed variation, as the prevalence of stag amongst treatments with a basin of stag of 0.2, for example, varies from 5% to 77% in period 8. This leads to the study of additional determinants of the prevalence of stag.

Other determinants: We study two sets of determinants of the prevalence of stag. First, we study two other determinants which are functions of the payoff parameters: the optimization premium (Battalio et al. (2001)), and hare's relative riskiness (Dubois et al. (2012)). Second, we study whether elements of the experimental design may have affected behavior.

Battalio et al. (2001) show that, in treatments when stag is not risk dominant, subjects are more likely to play stag when the incentives to choose a best response are smaller (in their terminology, that is when the *optimization premium* is small). Using our normalization, the optimization premium can be expressed as $\Lambda + \lambda$. Consistent with the results of Battalio et al. (2001), Table 4 shows in columns 3 to 6 that the optimization premium has a negative and mainly significant effect on the prevalence of stag.[8]

Dubois et al. (2012) show that the relative riskiness of hare (measured by the ratio of the payoff ranges of the two actions) is positively related to the prevalence of stag. Using our normalization, hare's relative riskiness can be expressed as $\frac{|1-\Lambda|}{1+\lambda}$. Consistent with the findings in Dubois et al. (2012), Table 4 shows in columns 3 and 4 that hare's relative riskiness has a positive effect on the prevalence of stag. However, this effect is only significant in period 8, and is not robust to the inclusion of controls for experimental design characteristics in columns 5 and 6.[9]

Part of the observed variation in behavior for a given basin of attraction may be explained by differences across previous articles in experimental design. For example some of the previous articles used the lottery method in an attempt to induce risk neutral preferences, and this may have affected behavior.[10] The use of the lottery method is not significantly related to the prevalence of stag in the specifications shown in columns 3 and 4 of Table 4, however, the effect is significant in period 8 when we include another control (see columns 5 and 6 in Table 4). This lack of robustness suggests that additional experiments are needed to study the impact of the lottery method on behavior in stag hunt games.

Different design implementations resulted also in differences in the number of times the same pair of subjects was expected to play together (as a function of the matching protocol and the total number of periods in a session). We find that this variable is statistically significant (see columns 5 and 6 in Table 4). This seems to explain part of the observed

[8] Given the small number of treatments in the metadata in which stag is risk dominant, we are not able to test whether the impact of the optimization premium differs when stag is risk dominant.

[9] We have also studied whether the normalization of payoffs to only two parameters discards information that is relevant to explaining behavior. We find that the constant and multiplicative coefficients used to normalize payoffs do not have a significant effect on behavior. Hence, our normalization of payoffs seems appropriate.

[10] Among the articles included in our metadata, Cooper et al. (1992), Straub (1995), and Duffy and Feltovich (2002) use the lottery method. Cooper et al. (1992) reports having also run experiments without the lottery method and finding that this method does not affect behavior. Studying whether the lottery method proposed by Roth and Malouf (1979) to induce risk neutral preferences affects behavior offers a window to study whether risk preferences are an important determinant of behavior. Finding that the lottery method affects behavior would suggest that risk preferences play an important role. Note, however, that there is disagreement on whether the lottery method affects risk preferences. Some studies find that it shifts behavior in line with theory while others find that it does not (see Harrison et al. (2013) and Kirchkamp et al. (2021) for recent examples of each case). As such, finding that this method does not affect behavior does not provide conclusive evidence that risk preferences do not matter, it could also be that the lottery method did not sufficiently affect risk preferences.

Table 5
Stag hunt game - row player's payoffs.

Original			Normalized		
	hare	stag		hare	stag
hare	60	T	hare	0	$1-\Lambda$
stag	S	90	stag	$-\lambda$	1

variation in behavior for the treatments with a basin of attraction equal to 0.2, with higher rates of stag in experiments in which the expected number of interactions for the same pair of subjects was higher.[11] Take for example, the games with lowest and highest prevalence of stag with a basin of attraction of 0.2. The former has an expected number of interactions of one and stag is selected 5% of the time, while the latter has an expected number of interactions of 10.71 and 65% play of stag. Based on the estimated probit model presented in the last column of Table 4, changing the number of interactions from 1 to 10.71 for the game with the lowest prevalence of stag would result in an increase of 32 percentage points in the prevalence of stag. If we also change the values of the optimization premium and hare's relative riskiness to those for the game with the high prevalence of stag, then the predicted prevalence of stag increases another 20 percentage points, to explain 52 percentage points of the observed 72 percentage point difference between the two games. That is, the determinants of behavior studied in this section can explain a large part of the observed difference in behavior among games with the same basin of attraction of stag.

It is important to note that the inclusion of these other possible determinants of play of stag into the analysis does not reduce the magnitude nor the significance of the impact of the size of the basin of attraction on the prevalence of stag (Table 4).

As already mentioned, the main limitation of the experiments conducted so far is that, although the original (non-normalized) payoffs are quite different across previous experiments, they hide a surprising degree of similarity with respect to the degree of strategic uncertainty. Most of them are in a narrow range for the basin of attraction of stag. Moreover, there are differences in experimental design that affected behavior and may confound the results. Therefore, to study the determinants of efficient coordination more systematically, we turn to a new experimental design that provides more variation in the variables of interest while keeping experimental methods constant.

4. The new experimental design

The main design innovation is to allow many more comparisons across parameters by presenting multiple stag hunt games simultaneously on the subjects' screen. More specifically, each session consists of 15 periods in which subjects participate anonymously through computers in the coordination games presented in Table 5.[12] Parameter T take values in the set {25, 45, 65, 85} and parameter S take values in the set {10, 20, 30, 40}. The relevant T and S for each stage games are known to subjects.[13] This results in 16 stag hunt games in each period.[14]

These games were displayed always in the same order, with S increasing across rows and T increasing across columns—see the decision screen in Figure 13 of the Appendix (which is available online).[15]

The set of possible values of T and S are selected to achieve two objectives. First, we want to have large and systematic variation in the parameters of the games. In particular we want to have large variations in the size of the basin of attraction of stag. The new experiments have the size of the basin of stag going from 0.091 to 0.765, with many intermediate values as shown in Table 6.[16] Second, we want to have many treatments for which stag is risk dominant so as to be able to study if that condition is sufficient for subjects to coordinate on the efficient equilibrium. Half the treatments have stag being risk dominant in the new experiment.

Subjects were randomly matched in each period to another subject, with the same pair not matched twice (perfect strangers). At the end of each period subjects received feedback on their own actions and the actions of the subject they were paired with for that period (see Figure 14 in the Appendix). Subjects could access information on the past outcomes of the games in which they had participated (see Figure 15 in the Appendix).

[11] Note that given the limited number of studies, we cannot disentangle the role of the rematching protocol from the expected number of interactions.

[12] The actions were simply described as "1" and "2" in the experiment.

[13] In terms of normalized payoffs, the 16 games have Λ in the set $\{\frac{1}{6}, \frac{5}{6}, \frac{3}{2}, \frac{13}{6}\}$ and λ in the set $\{\frac{2}{3}, 1, \frac{4}{3}, \frac{5}{3}\}$.

[14] This experimental design is related to the strategy method as subjects need to simultaneously decide what they would do under alternative circumstances. A possible issue with the strategy method is that it may affect behavior. Hoffman et al. (1998), Gueth et al. (1998), and Brosig et al. (2003) report evidence that the strategy method may affect behavior while Brandts and Charness (2000) find that it does not. See Brandts and Charness (2011) for a review of the literature. We address this issue in Sections 7 and 8.

[15] Ex-ante it is difficult to know if the order of presentation should help or hinder subjects from making choices consistent with some rule. However, the order of presentation to subjects is designed such that only picking rows or columns would not be consistent with a basin of attraction based choice rule. In the end, as indicated in the learning model (see Section 8), it seems to make it more challenging (noisier).

[16] These 16 stag hunt games also result in 11 values of the optimization premium (going from 0.83 to 3.83) and 15 values of hare's relative riskiness (going from 0.063 to 0.7).

Table 6
Size of basin of attraction of stag.

Λ	λ			
	2/3	1	4/3	5/3
1/6	0.2	0.143	0.111	0.091
5/6	**0.556**	0.455	0.385	0.333
3/2	**0.692**	**0.6**	**0.529**	0.474
13/6	**0.765**	**0.684**	**0.619**	**0.565**

Note: bold font denotes stag is risk dominant.

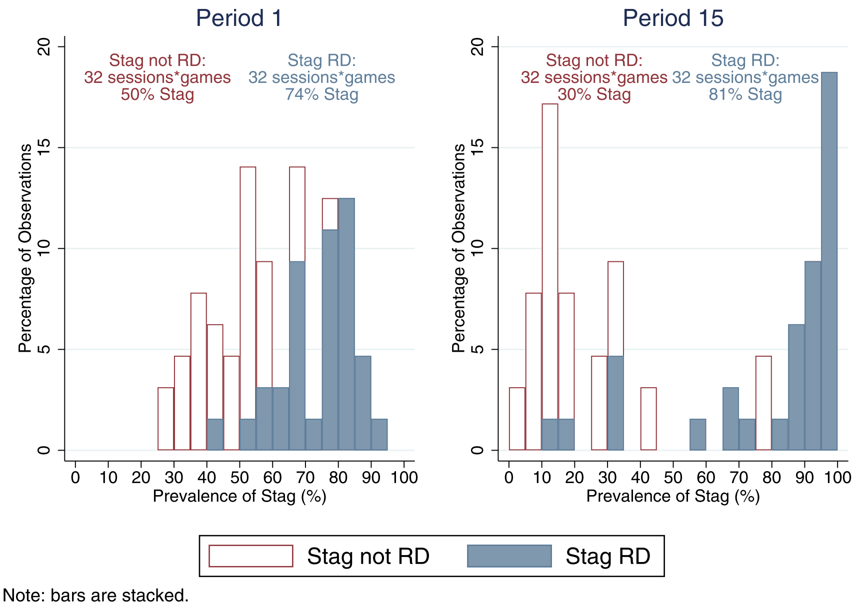


Fig. 3. *Baseline*: distribution of prevalence of stag by session and game.

Subjects were paid based on the points earned in one randomly chosen game for one randomly chosen period. We have two main treatments, *Baseline* and *Lottery*, which differ by how points are transformed into payments. In *Baseline*, points are exchanged into dollars at the rate of $35 per 100 points. In *Lottery*, points are the chances of receiving $35. In addition, in both treatments, subjects are paid a $5 participation fee and a $5 show-up fee. Note that for any given outcome of the game, the expected payoff is equal across the two treatments, thus facilitating comparisons.

The experiment was programmed using z-Tree (Fischbacher, 2007). We conducted four experimental sessions for each of these two treatments with a total of 140 subjects. See Table 11 in the Appendix for the number of subjects per session and treatment. The experimental sessions lasted less than an hour. The subjects were Brown University undergraduates recruited through advertisement in university web pages, leaflets, and signs posted on campus. Subjects earned $34.35 on average, with a minimum of $10 and a maximum of $45, including the participation fee and the show-up fee of $10.[17]

5. Aggregate analysis of the determinants of efficient behavior

We start this section by focusing on the *Baseline* treatment and studying the conditions under which subjects choose the efficient action (stag).

Payoff dominance: Averaging across games, 61% of subjects chose stag in period 1 and 51% in period 15. Hence, it is not the case that subject coordinate on stag regardless of the payoff matrix. Fig. 3 shows the distribution of the prevalence of stag for each session and game. We see that the prevalence of stag goes from 0% to 100% in period 15, showing again that people do not necessarily coordinate on the payoff dominant equilibrium. A similar conclusion is reached if we focus on the average behavior by game. Note that in period 1, a majority of subjects chooses stag in only 11 of the 16 games and this number is reduced to 7 in period 15—see Fig. 4.

[17] The minimum of $10 and the maximum of $45 were both reached in the *Lottery* treatment. In the *Baseline* treatment the minimum and maximum earnings were $17 and $41.50.

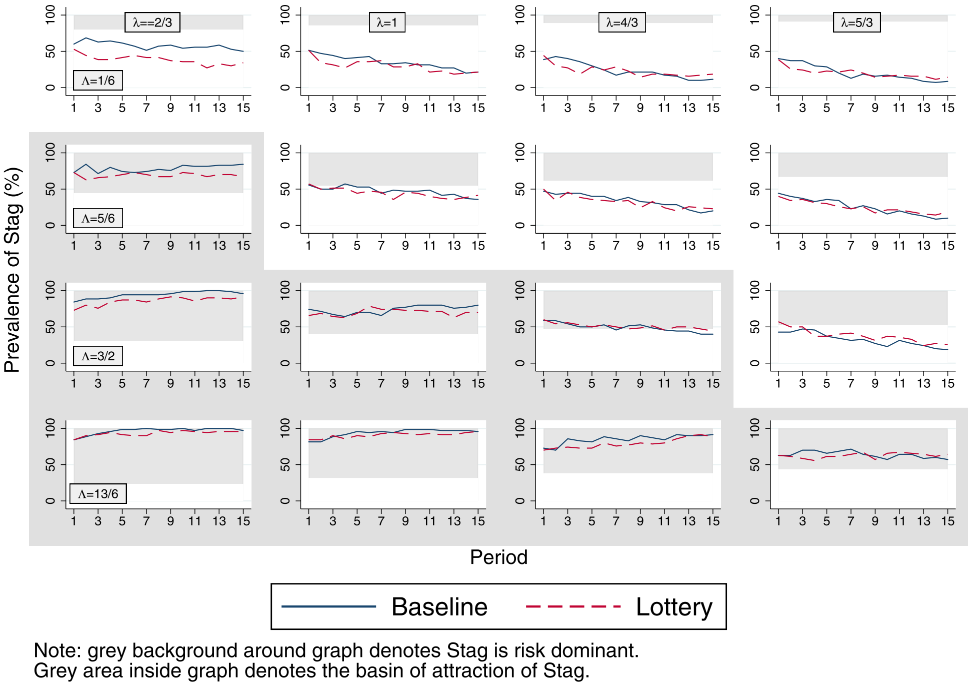


Fig. 4. Evolution of behavior in *Baseline* and *Lottery*.

Risk dominance: It is the case that the prevalence of stag is significantly higher in games in which it is risk dominant—see Fig. 3. In period 1, there is 50% of stag when it is not risk dominant versus 74% when it is risk dominant (p-value $<$ 0.01).[18] In period 15, the prevalence of stag is 30% and 81% respectively (p-value $<$ 0.01). However, even after controlling for risk dominance there are large variations in the prevalence of stag as shown in Fig. 3. Note that for treatments in which stag is risk dominant, the prevalence of stag goes from 15% to 100% in period 15. For treatments in which stag is not risk dominant, the prevalence of stag goes from less than 5% to 75%. There is a large variation of behavior that remains to be explained.

Strategic uncertainty (beyond risk dominance): We study now whether, beyond risk dominance, the prevalence of stag is correlated to the robustness of the efficient equilibrium to strategic uncertainty, as measured by the size of the basin of attraction. Fig. 5 shows the prevalence of stag as a function of the basin of attraction for periods 1, 8, and 15. As with the metadata, the correlation between the size of the basin and the prevalence of stag is positive and increases with experience in the new experiment as well. The first 3 columns in Table 7 show that the size of the basin of attraction of stag has a small effect on behavior if stag is not risk dominant, while it has a large and significant effect if stag is risk dominant. This is consistent with the findings regarding the effect of the size of the basin of attraction of Always Defect on cooperation in repeated games—see Dal Bó and Fréchette (2018).

Other determinants: As we did in the meta-analysis, we study now if the prevalence of stag is affected by the optimization premium (Battalio et al. (2001)) and hare's relative riskiness (Dubois et al. (2012)).

The last 3 columns in Table 7 show that the optimization premium has a negative and significant effect on the prevalence of stag: the larger the incentives to best respond to the actions of others, the lower the share of subjects attempting to coordinate on the efficient equilibrium. This is consistent with what we found in the meta-analysis and with Battalio et al. (2001). Note that this effect is similar regardless of whether stag is risk dominant or not. The significant negative effect of the optimization premium when stag is risk dominant is somewhat surprising, as one would expect that a greater optimization premium would facilitate coordination on the efficient equilibrium in that case.

Consistent with Dubois et al. (2012), Table 7 shows that hare's relative riskiness has a positive impact on the prevalence of stag (the impact is statistically significant in periods 8 and 15).

We next study the relative importance of the different determinants of behavior that we have studied in this section. We take the estimates from column 6 in Table 7 and measure the predicted change in the prevalence of stag when we allow one variable to go from the minimum to the maximum of its values in the experiment while keeping the other variables

18 We compute significance from a probit analysis at the individual level clustering standard errors by session (see Fréchette (2012) for a discussion).

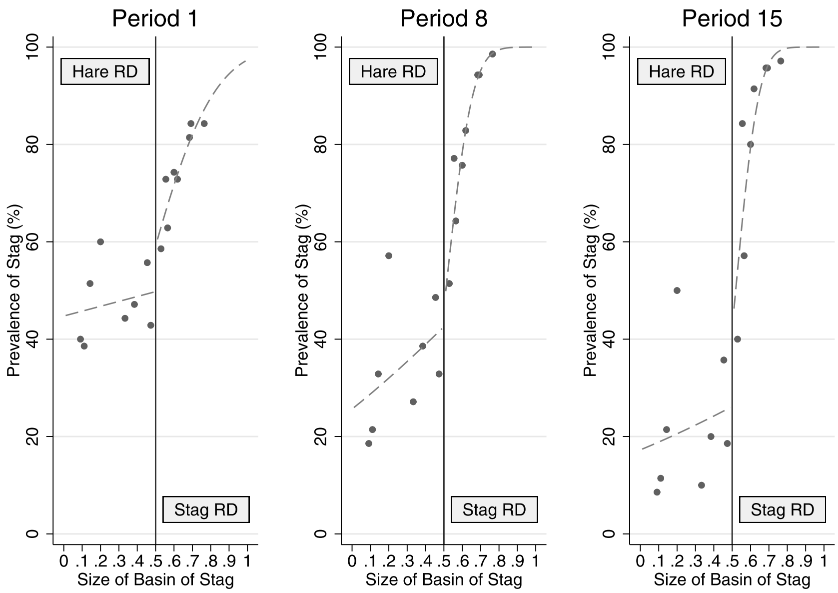


Fig. 5. Relation between the prevalence of stag and the size of the basin.

Table 7
Baseline: determinants of stag (probit analysis - marginal effects).

	Period 1	Period 8	Period 15	Period 1	Period 8	Period 15
Stag RD (d)	-0.48***	-0.93***	-0.97***	-0.36**	-0.70**	-0.40*
	(0.108)	(0.077)	(0.034)	(0.155)	(0.312)	(0.237)
RD × Basin	1.30***	3.33***	3.99***	1.18***	2.70***	2.82***
	(0.212)	(0.679)	(0.480)	(0.185)	(0.813)	(0.197)
Not RD × Basin	0.10	0.35*	0.24	0.54***	1.30***	1.74***
	(0.086)	(0.200)	(0.181)	(0.153)	(0.212)	(0.172)
Not RD × OP				-0.12***	-0.17***	-0.23**
				(0.039)	(0.010)	(0.093)
RD × OP				-0.06**	-0.12***	-0.23**
				(0.026)	(0.041)	(0.114)
Hare's RR				0.08	0.42***	0.78***
				(0.109)	(0.131)	(0.242)
Observations	1120	1120	1120	1120	1120	1120

(d) denotes dummy variable, effect of change from 0 to 1 is reported. Basin denotes the size of the basin of stag. OP denotes the optimization premium ($\Lambda + \lambda$). Hare's RR denotes hare's relative riskiness $\left(\frac{|1-\Lambda|}{1+\lambda}\right)$.
Standard errors clustered at the session level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

at their average value. We find that the shift of the size of the basin of attraction of stag from its lowest to its highest value increases the predicted prevalence of stag by 93 percentage points, while the change is 65 percentage points for the optimization premium and 17 percentage points for hare's relative riskiness.

The calculations of the different determinants of behavior studied here are done assuming that subjects are risk neutral, which may not be the case.[19] The *Lottery* treatment allows us to study whether our reliance on the assumption of risk neutrality may be problematic. Finding that the lottery method affects behavior would suggest that risk preferences are a determinant of behavior in stag hunt games. However, given the disagreement on whether this method affects risk preferences (see footnote 10), finding that the lottery method does not affect behavior could be attributed to either risk preferences not being an important determinant of behavior or the lottery method not affecting risk preferences in our experiment.

Fig. 4 displays the evolution of the prevalence of stag for all games for the *Lottery* treatment in addition to the *Baseline* treatment. It is clear that both the levels and evolution of behavior are very similar in these two treatments. In period

[19] Büyükboyacı (2014) elicits risk preferences and finds no correlation between those and behavior in stag hunt games. For related games, Cason et al. (2012) find no significant relationship between elicited risk preferences and behavior in minimum effort games, while Heinemann et al. (2009) find a relationship between elicited risk preferences and behavior in threshold public goods games.

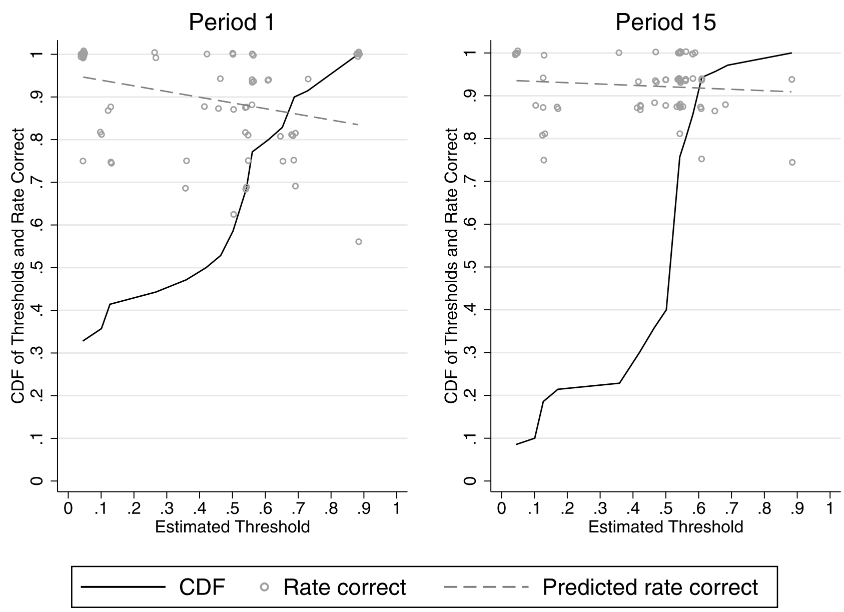


Fig. 6. Distribution and fit of estimated thresholds.

1, behavior is significantly different at the 5% level in three out of 16 games, and there are no significant differences by period 15. Moreover, as shown in Appendix Table 13, the results on the impact of risk dominance and the size of the basin of attraction on behavior are robust to including the *Lottery* treatment with no significant differences between *Lottery* and *Baseline*. As discussed above, this result is consistent with risk preference not playing an important role in this experiment but it is not conclusive, as we have not directly measured whether the lottery method affects risk preferences in our experiment.

6. Analysis of individual behavior

The new experimental design allows us to study the determinants of behavior in stag hunt games by looking at the data at the individual level. In particular, we study if subjects' choices appear consistent with decision rules based on the size of the basin of attraction. We also study the evolution of such decision rules over time and differences across subjects.

We assume that each subject follows a threshold decision rule such that the subject chooses stag if the basin of attraction of stag for that game is greater than the threshold. For each subject and period, we compute the threshold that best corresponds to the subject's choices in that period across the 16 games: that is the threshold that minimizes the number of errors.[20] Given the finiteness of the data (we observe behavior in 16 games with different basins of attraction), the best we can do is to estimate in which interval of possible values a threshold may fall. For example, if we observe a subject playing stag for all games with basin of attraction equal or greater than 0.2 in the experiment, then we know that the threshold that best explain this behavior must be between 0.2 and the next lower available basin of attraction in the experiment (0.143). For a simple presentation of results, we identify each possible interval of thresholds with the middle of each interval.[21]

Also note that there may be more than one threshold that minimizes the number of errors (this happens when the observed realization of behavior is non-monotonic in the basin of attraction). In 70% of the cases there is a unique estimated threshold, two thresholds in 24% of the cases, and three thresholds in the remaining cases. For simplicity, in what follows we choose the minimum threshold when multiple ones are available for a subject; results are similar if we use the maximum one. The fit of the estimated threshold is tight as the estimated thresholds explain behavior in more than 90% of the actions in every period.

Fig. 6 displays the CDF of the estimated thresholds for periods 1 (left panel) and 15 (right panel). This figure also shows the predictive power (*rate correct*) of the estimated threshold based rule for each subject (the ratio of decisions that are explained by the estimated rule). The hollow circles indicating *rate correct* are *jittered* to help distinguish the coordinates with multiple points. The high average values of *rate correct* displayed in Fig. 6 show that a threshold based choice rule accounts for a significant fraction of choices for most subjects (the average value of *rate correct* is 0.9 for period 1 and 0.93

[20] As in Rankin et al. (2000) who also estimate threshold decision rules in stag hunt games, Levine and Palfrey (2007) who estimate threshold decisions rules in voting turnout, and Fréchette et al. (2020) who estimate receiver's decision rules in a communication game.

[21] Results are very similar if we focus on either the maximum or the minimum of the intervals. The size of the interval is on average 0.06.

for period 15). The grey lines in the figure indicate the linear fit of *rate correct* with respect to thresholds. Interestingly, the quality of the fit does not noticeably increase with experience or across the estimated thresholds.

Regarding the estimated thresholds, we observe a large distribution of thresholds in period 1 with the most prevalent one being the lowest possible threshold (play stag in every game), and the second most prevalent one being 0.54. We observe a convergence towards middle thresholds with the most popular threshold being 0.54 in period 15.

To help describe the patterns of behavior the figure reveals, it is useful to divide thresholds into low ($< \frac{1}{3}$), medium (between $\frac{1}{3}$ and $\frac{2}{3}$), and high ($\geq \frac{2}{3}$). In period 1, most thresholds are low (44% of cases; i.e. subjects mostly select stag), followed by thresholds in the middle tercile (39%). However, by period 15, medium thresholds compose the vast majority (74%). This increase in popularity of the medium thresholds comes mostly from subjects abandoning low thresholds (29% of the data moves from a low to a middle threshold). This includes an important decrease in subjects whose choice rule is best described by the lowest possible threshold (play stag in every game), going from 33% of subjects in period 1 to below 9% in period 15. However, a movement in the opposite direction also occurs: 10% of the subjects go from high thresholds to medium thresholds.

Consistent with this movement toward central thresholds, there is an increase in subjects best described as following risk dominance (those with one of their estimated thresholds intervals including 0.5). Only 9% of subjects can be best described as following risk dominance in period 1, and this number increases to 16% by period 15. However, this does not seem to be a very stable choice pattern, as few subjects fit the risk dominant choice pattern repeatedly. In fact, the one subject that most often chooses in a way most consistent with risk dominance does it in 12 of the 15 periods (80%). No other subjects chooses in a way most consistent with risk dominance in more than 50% of the periods.[22]

7. Is the new experimental design neutral?

The design novelty in the new experiments presented in this paper is to let subjects participate in several coordination games simultaneously in each period. This allows us to gather data on a greater number of games than it would be possible if subjects only played one game per period and it also allows us to study behavior across games at the individual level. But is this design neutral? Is it possible that behavior is affected by subjects playing several games simultaneously?

To answer these questions we present results from two additional treatments that differ from *Baseline* in that subjects play only one stag hunt game in every period. One of the treatments considers the stag hunt game with $\Lambda = \frac{3}{2}$ and $\lambda = 1$ (in the non-normalized payoffs seen by the subjects that is $T = 45$ and $S = 30$), and the other treatment considers the stag hunt game with $\Lambda = \frac{5}{6}$ and $\lambda = \frac{4}{3}$ ($T = 65$ and $S = 20$). These two treatments, *One Game* $\frac{3}{2}$&1 and *One Game* $\frac{5}{6}$&$\frac{4}{3}$ allow us to compare behavior with the same games in the *Baseline* treatment.[23]

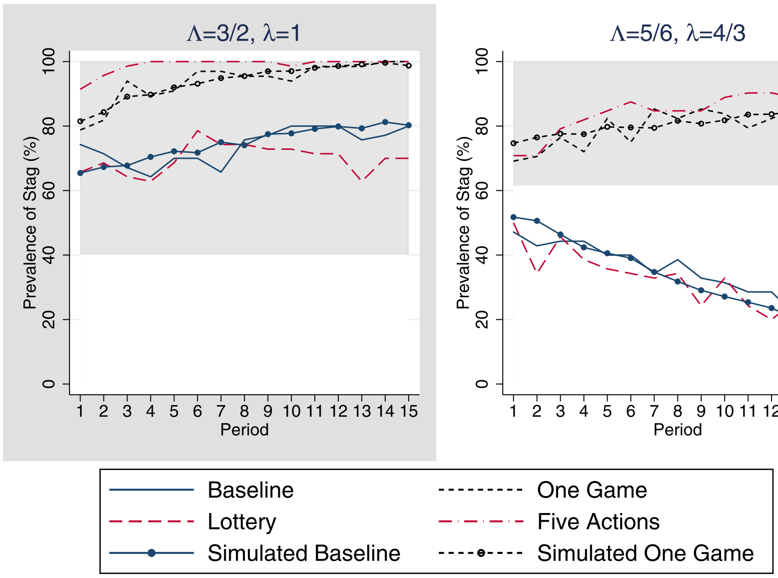
We conducted four experimental sessions for each of these treatments with a total of 134 subjects. See Table 11 in the Appendix for the number of subjects per session and treatment. The experimental sessions lasted less than an hour. Subjects earned \$36.20 on average, with a minimum of \$17 and a maximum of \$41.5, including the participation and the show-up fee of \$10.

As can be seen in Fig. 7, behavior is quite different between the *Baseline* treatment and the *One Game* treatments. For $\Lambda = \frac{3}{2}$ and $\lambda = 1$, the prevalence of stag is greater under *One Game* than under the *Baseline* in period 1 (but this difference is not statistically significant at the 10% level). The difference increases after the first two periods and remains statistically significant at the 1% level until the end. For $\Lambda = \frac{5}{6}$ and $\lambda = \frac{4}{3}$, the prevalence of stag is greater under *One Game* than under the *Baseline* in period 1 (p-value < 0.05). The difference increases after the first period and becomes statistically significant at the 1% level until the end. We find a similar pattern if we compare the *One Game* and the *Lottery* treatments.

The differences between our *One Game* treatments and the *Baseline* are consistent with the differences in behavior between the *Baseline* and prior studies (in which subjects also played one game at a time) for intermediate values of the basin of attraction of stag. As such, the comparison across treatments and papers shows that choice frequencies in stag hunt games may depend on whether subjects play one game in isolation or several games simultaneously. However, regardless of this difference, the comparative statics are robust: the prevalence of stag increases with its robustness to strategic uncertainty, as measured by the size of its basin of attraction, whether the subjects only play one game at a time or many. This is also the case for the impact of the optimization premium and, to a lesser degree, for the impact of hare's relative riskiness. As it is not clear whether playing each game in isolation or playing several simultaneous games is more realistic or relevant for applications, we believe that focusing on results that are robust to the details of the experimental design is of importance.

[22] In period 1, only 3% of the subjects choose actions that are perfectly consistent with risk dominance, in period 15 no subject does that.

[23] We also considered an alternative design to test the neutrality of having subjects participate in 16 simultaneous games: have two treatments in which subjects participate in two different, but overlapping, subsets of the 16 games considered in the *Baseline* treatment. This would have allowed us to directly measure spillover effects across games by comparing behavior in the overlapping games. We opted for the design presented in this section as it allows a direct comparison of the *Baseline* treatment with the more standard design in which subjects participate in one game at a time.



Note: grey background denotes Stag is risk dominant.
Grey area inside graph denotes the basin of attraction of Stag.

Fig. 7. Evolution of behavior in all treatments.

8. What explains the non-neutrality of the new design?

In this section, we investigate some possible reasons for the differences in behavior between *One Game* treatments and the two treatments with several games per period introduced in this paper (*Baseline* and *Lottery* treatments). In particular we consider: 1) the effect of neighboring games when they are presented simultaneously (spillover effects); 2) the effect of the number of possible strategies; and 3) the effect of the number of simultaneous games on the level of noise in decision making.

Spillover effects: Playing several simultaneous games may affect behavior due to spillover effects across games when played simultaneously.[24] We show here that such spillover effects exist in the *Baseline* and *Lottery*, but that they cannot explain the observed difference in behavior with the *One Game* treatments.

We study the existence of spillover effects in the *Baseline* and *Lottery* treatments by using the characteristics, in terms of the size of the basin of attraction, of neighboring games on the screen with 16 games to explain behavior. We construct two measures of the characteristics of the neighbors. First, we focus on the four neighbors (left, right, top, and bottom) of each game and calculate their average size of the basin of attraction of stag.[25] Second, we focus on the eight immediate neighbors surrounding the game in consideration.[26]

As shown in Table 8, regardless of the definition of neighbors, it is the case that the larger the average size of the basin of attraction of stag for the neighbors, the greater the share of subjects choosing stag. The effect is statistically significant for the first period but not for the last one. Note, however, that the magnitude of the effects does not decrease significantly as subjects gain experience.

While finding these spillover effects is of interest on it own, as it shows that behavior in coordination games may depend on elements beyond the payoff structure of that coordination game considered in isolation, they cannot explain the difference in behavior between the multiple games and one game treatments. The reason is that the two games we use to compare the two experimental designs do not have neighboring games in the *Baseline* and *Lottery* treatments with drastically different basins of attraction than their own: the average basin of attraction of the eight neighbors is 0.59 (as opposed to 0.6) for the game with $\Lambda = \frac{3}{2}$ and $\lambda = 1$ and 0.34 (vs 0.385) for the game with $\Lambda = \frac{5}{6}$ and $\lambda = \frac{4}{3}$. In fact, if we compute the predicted rates of stag replacing the value of the neighboring games with the basin of the actual game (thus nullifying the spillover effect), the effect is very modest and the predicted prevalence of stag is far below what is observed

[24] Bednar and Page (2007), Huck et al. (2011), and Bednar et al. (2012) study simultaneous spillover effects, and Cooper and Kagel (2008), Cooper and Kagel (2009), and Rick and Weber (2010) study sequential spillover effects, or order effects. Relatedly, Rankin et al. (2000) and Kendall (2020) study behavior in a sequence of different stag hunt games. Rankin et al. (2000) find a high prevalence of stag while Kendall (2020) does not. The difference may be due to the fact that the latter focuses on games where stag is not risk dominant while the former does not. Van Huyck and Battalio (2002) use a similar experimental design to study behavior in bargaining games.

[25] For games on the border of the combinations of S and T, this may consist of the average of only two or three numbers.

[26] For games on the border of the combinations of S and T, this may consist of the average of only three or five numbers.

Table 8
Effect of neighbors' size of basin of attraction on stag (marginal effects from probit - *Baseline* and *Lottery*).

	Period 1	Period 15	Period 1	Period 15
RD (d)	-0.40***	-0.93***	-0.40***	-0.94***
	(0.061)	(0.104)	(0.063)	(0.088)
RD × Basin	0.98***	3.14***	1.08***	3.57***
	(0.107)	(1.074)	(0.120)	(0.988)
Not RD × Basin	0.04	-0.00	0.13*	0.27**
	(0.089)	(0.190)	(0.067)	(0.119)
4 Neighbors' Basin	0.25***	0.58		
	(0.072)	(0.488)		
8 Neighbors' Basin			0.16**	0.15
			(0.072)	(0.408)
Observations	2240	2240	2240	2240

Marginal effects; Standard errors clustered at the session level in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9
Five action games - row player's payoffs.

Λ=3/2, λ=1 (T=45, S=30)

	hare	A	B	C	stag
hare	60	56	53	49	45
A	53	68	64	60	56
B	45	60	75	71	68
C	38	53	68	83	79
stag	30	45	60	75	90

Λ=5/6, λ=4/3 (T=65, S=20)

	hare	A	B	C	stag
hare	60	61	63	64	65
A	50	68	69	70	71
B	40	58	75	76	78
C	30	48	65	83	84
stag	20	38	55	73	90

for the *One Game* treatment.[27] While there are spillover effects, they cannot explain the difference in behavior between treatments with one game per period and those with several games per period.

The number of strategies: An alternative candidate to explain the observed difference between *One Game* and *Baseline* and *Lottery* treatments comes from the literature on the determinants of cooperation in infinitely repeated games experiments. As we pointed out in the introduction, the basin of attraction has been found to predict behavior in infinitely repeated prisoner's dilemma games. Interestingly, the relation between the basin of attraction of the cooperative strategy and the rate of cooperation found in that literature is very similar to the one observed for stag and its basin in *Baseline* and *Lottery*: the prevalence of the efficient behavior is low for intermediate values of the basin of attractions—see Dal Bó and Fréchette (2018).

One possible factor that could explain why the relation is similar for the *Baseline* and *Lottery* treatments and infinitely repeated prisoner's dilemma games, but different for the *One Game* treatments, is that the impact of strategic uncertainty may be mediated by the number of possible strategies the subject must choose from. There are an infinite number of possible strategies in an infinitely repeated game, and there are 16^2 possible combinations of choices (or strategies) in a period of the *Baseline* and *Lottery* treatments as a subject needs to choose between two action in 16 games. If the number of possible strategies affects coordination, we expect that changing the number of possible actions in a coordination game would affect behavior. To explore this possibility, we conduct two additional treatments using the *One Game* paradigm, but with a more complex game in that each game has five actions. Table 9 represents the payoff matrices for a game with five actions which extends the payoff matrix used in the *One Game* $\frac{3}{2}$&1 and *One Game* $\frac{5}{6}$&$\frac{4}{3}$ treatments.

Our aim was to make the payoffs of the five action games as close to those of the corresponding games with 2 actions. As presented in Table 9, hare and stag are placed in each corner of the table so that the salience of hare and stag is only minimally affected.[28] For these additional treatments, *Five Actions* $\frac{3}{2}$&1 and *Five Actions* $\frac{5}{6}$&$\frac{4}{3}$, the experimental design differs from that of the *One Game* treatments only in the different payoff matrices.

We conducted four experimental sessions for each of these two additional treatments with a total of 142 subjects. See Table 11 in the Appendix for the number of subjects per session and treatment. The experimental sessions lasted less than an hour. Subjects earned \$38.56 on average, with a minimum of \$17 and a maximum of \$41.5, including the participation and the show-up fee of \$10.

[27] If we focus on the probit analysis presented in the second column of Table 8 (which estimates the strongest spillover effects), replacing the average size of the basin of attraction of the neighbors with the one from the actual game changes the predicted prevalence of stag by less than 2 percentage points. Moreover, the predicted prevalence is far from the observed one under the *One Game* treatment even if we replace the average size of the basin of attraction of the neighbors with the maximum one in the experiment.

[28] The actions were simply described as "1" to "5" in the experiment.

Table 10
Learning model estimates.

	$\Lambda = \frac{3}{2}$&$\lambda = 1$		$\Lambda = \frac{5}{6}$&$\lambda = \frac{4}{3}$	
	Base.	*One*	*Base.*	*One*
σ	16.08	7.58	8.59	2.71
	(0.103)	(0.096)	(0.077)	(0.234)
β_0^{Stag}	1.36	3.78	2.79	34.67
	(0.048)	(0.030)	(0.110)	(5.151)
β_0^{Hare}	1.57	4.61	4.65	68.22
	0.050	(0.055)	(0.114)	(8.843)
θ	0.9			
	(0.001)			

Standard errors clustered at the session level in parentheses.

Fig. 7 shows the evolution of behavior in the *One Game* and *Five Actions* treatments and the corresponding games in the *Baseline* and *Lottery* treatments. For the games with five actions, the prevalence of stag is computed by the relative frequency over hare and stag.[29] It is clearly the case that the additional treatments with five actions reveal a pattern of behavior that is similar to the one observed for the *One Game* treatments with two actions. This suggests that the number of possible strategies the subject must choose from is unlikely to be the driving force for the difference between our *Baseline* and *One Game* treatments.

The number of simultaneous games and noise: While the subjects must make one choice per period in the *One Game* and *Five Actions* treatments, they must make choices in 16 games per period in the *Baseline* and *Lottery* treatments. The requirement to make more simultaneous choices and to absorb feedback about 16 games, rather than one, could plausibly increase *noise* in choices. To explore this possibility, we consider whether a simple belief based model that allows for different levels of noise can reproduce the differences observed, and show that the complexity that arises from playing 16 simultaneous games increases noise affecting both the level and evolution of behavior.[30]

Belief based learning models have been estimated in the context of coordination games, see for instance Crawford (1995), Cheung and Friedman (1997), and Battalio et al. (2001). The specific belief based model we consider is a version of the ones estimated in Fréchette (2009), Dal Bó and Fréchette (2011), and Embrey et al. (2017).[31] The model specifies that agents maximize expected utility, which is determined by material payoffs, beliefs that the other player will select each of her or his actions, and an unobserved idiosyncratic term. The beliefs are determined as the ratio of weights attached to each choice: $\beta_{i,t}^{Stag}$ and $\beta_{i,t}^{Hare}$. The unobserved idiosyncratic shock $\epsilon_{i,t}$ is distributed according to an extreme value distribution with parameter σ determining its variance. The weight on each choice updates as a function of experience. Agents put more weight on choices they observe other players using. Specifically, $\beta_{i,t+1}^k = \theta\beta_{i,t}^k + 1\{a_{-i,t} = k\}$ where a is the action taken and $\theta \in [0, 1]$ controls the importance of past weights relative to the last observed choice. At one extreme, $\theta = 0$, the model corresponds to Cournot learning, and the other, $\theta = 1$, the model allows for fictitious play. Also, beliefs at any period are more correlated with the beliefs in the first period when the $\beta_{i,0}^k$'s are larger.

For the estimation we use the *Baseline* and *One Game* treatment games $\Lambda = \frac{3}{2}$&$\lambda = 1$ and $\Lambda = \frac{5}{6}$&$\lambda = \frac{4}{3}$.[32] The weights and variance of the error term are estimated separately by game and treatment. For simplicity, θ is assumed to be the same in both games and both treatments. Hence, 13 parameters are estimated, four estimates for β_0^{Stag}, β_0^{Hare}, and σ; and one for θ.

Results of the estimation are presented in Table 10. The estimated noise in decision making (σ) is smaller in the *One Game* treatment than in the *Baseline* treatment for both games. This difference is jointly significant at the 10%, however, when doing the test for specific games, it is not significant for $\Lambda = \frac{5}{6}$&$\lambda = \frac{4}{3}$ (and it is at the 5% level for the other). These estimates suggest that subjects are more responsive to expected utility when they participate in one game at a time. The differences in estimated β's on the other hand are not statistically different when considered jointly (and neither when considering specific games). However, the estimates are smaller in the *Baseline* treatment than in the *One Game* treatment for both games. As such, the estimates of the β's are consistent with subjects having stronger priors when participating in one game at a time. Furthermore, for both games, the Baseline results imply beliefs closer to 50-50 when subjects participate in several games simultaneously (specifically a prior on Stag of 54% and 55% in the *Baseline* treatment versus 63% and 66% in the *One Game* treatment). Both stronger priors in the *One Game* treatment and priors closer to 50% in the *Baseline* are consistent with subjects expecting others to be noisier in the *Baseline*.

[29] Only a minority of subjects chose actions besides stag and hare. The distribution of actions is shown in Figure 12 in the Appendix.

[30] In the context of infinitely repeated prisoner's dilemma, where a coordination element is present, Proto et al. (2021) show that noise plays a role in the evolution of play.

[31] Dal Bó and Fréchette (2011) and Embrey et al. (2017) estimate parameters for each subject. Here, as in Fréchette (2009), where we have few data points per subject and treatment, we pool the data and account for heterogeneity by allowing for within session correlation in the error terms.

[32] We do not include data from the *Lottery* so that we have a similar number of observations for treatments with several games per period and one game per period.

Of particular interest is seeing if the estimated initial weights and noise levels can generate comparative statics similar to what is observed in the data across the *Baseline* and *One Game* treatments. For this we simulate the evolution of behavior in 100 sessions per game and treatment based on the estimated parameters reported in Table 10. Fig. 7 presents the average observed prevalence of stag in these simulations. As the figure makes clear, the simple model with treatment-specific noise and priors can capture the small increase in rates of stag in both treatments for the game in the left panel ($\Lambda = \frac{3}{2} \& \lambda = 1$), while generating diverging trends across treatments for the game in the right panel ($\Lambda = \frac{5}{6} \& \lambda = \frac{4}{3}$). For the game in the left panel, the initial beliefs fall in the basin of attraction of stag for both treatments. As such it is a best response for subjects to play stag. Given the difference in noise across treatments, the prevalence of stag is greater for the treatment with lower noise (*One Game*) but in both treatments subjects tend to update priors in favor of stag increasing the incentives to choose it as they gain experience. For the game in the right panel of Fig. 7 the beliefs fall in the basin of attraction of stag for the *One Game* treatment and in the basin of attraction of hare for the *Baseline* treatment. In addition, behavior is closer to 50% in the latter. The initial beliefs together with the initial distribution of behavior put the two treatments on different trajectories.

9. Conclusions

We use the metadata from previous experiments and data from a new experiment to study the determinants of efficient behavior in stag hunt games. We find that subjects do not necessarily play the efficient action (stag), stressing the importance of strategic uncertainty in stag hunt games. Consistent with that idea, we find that subjects are significantly more likely to choose stag when it is risk dominant. However, risk dominance leaves much variation of behavior to be explained. We show that other determinants of behavior help explain this variation. We find that the optimization premium and hare's relative riskiness are also part of the explanation. Tightly connected with the idea that strategic uncertainty matters, we find that the prevalence of stag increases with the robustness of the efficient equilibrium to strategic uncertainty: as the basin of stag increases, the prevalence of stag tends to increase. The importance of the size of the basin of attraction as a determinant of behavior is validated by exploiting within-subject variation of behavior in our new experiment. Most subjects can be described as following a monotonic decision rule (play stag if its basin of attraction is greater than a given number).

However, the exact relation between the size of the basin of attraction of stag and its prevalence depends on the number of games subjects play in a given period. For intermediate values of the size of the basin of attraction of stag, the prevalence of stag is higher when subjects play only one game per period than when they play 16 games simultaneously as in our new experimental design. We provide evidence that this difference can be explained by subjects displaying greater noise in decision making when they are playing several games simultaneously.

Our findings reveal a strong and stable qualitative relationship that is independent of the details of the experimental design: as the size of the basin of attraction of stag increases and the optimization premium decreases, subjects become more likely to choose the efficient action.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.geb.2021.08.010.

References

Battalio, R., Samuelson, L., Van Huyck, J., 2001. Optimization incentives and coordination failure in laboratory stag hunt games. Econometrica 69 (3), 749–764.
Bednar, J., Chen, Y., Liu, T.X., Page, P., 2012. Behavioral spillovers and cognitive load in multiple games: an experimental study. Games Econ. Behav. 74 (1), 12–31.
Bednar, J., Page, S.E., 2007. Can game(s) theory explain culture?: The emergence of cultural behavior in multiple games. Ration. Soc. 19 (1), 65–97.
Blonski, M., Spagnolo, G., 2015. Prisoners' other dilemma. Int. J. Game Theory 44 (1), 61–81.
Brandts, J., Charness, G., 2000. Hot vs. cold: sequential responses and preference stability in experimental games. Exp. Econ. 2, 227–238.
Brandts, J., Charness, G., 2011. The strategy versus the direct-response method: a first survey of experimental comparisons. Exp. Econ. 14, 375–398.
Brosig, J., Weimann, J., Yang, C.-L., 2003. The hot versus cold effect in a simple bargaining experiment. Exp. Econ. 6, 75–90.
Büyükboyacı, M., 2014. Risk attitudes and the stag-hunt game. Econ. Lett. 124 (3), 323–325.
Calford, E., Oprea, R., 2017. Continuity, inertia, and strategic uncertainty: a test of the theory of continuous time games. Econometrica 85 (3), 915–935.
Cason, T.N., Savikhin, A.C., Sheremeta, R.M., 2012. Behavioral spillovers in coordination games. Eur. Econ. Rev. 56 (2), 233–245.
Castillo, M., Dianat, A., 2018. Strategic uncertainty and equilibrium selection in stable matching mechanisms: experimental evidence. Working Paper.
Charness, G., 2000. Self-serving cheap talk: a test of Aumann's conjecture. Games Econ. Behav. 33 (2), 177–194.
Cheung, Y.-W., Friedman, D., 1997. Individual learning in normal form games: some laboratory results. Games Econ. Behav. 19 (1), 46–76.
Clark, K., Kay, S., Sefton, M., 2001. When are Nash equilibria self-enforcing? An experimental analysis. Int. J. Game Theory 29 (4), 495–515.
Cooper, R.W., DeJong, D.V., Forsythe, R., Ross, T.W., 1990. Selection criteria in coordination games: some experimental results. Am. Econ. Rev. 80 (1), 218–233.
Cooper, R.W., DeJong, D.V., Forsythe, R., Ross, T.W., 1992. Communication in coordination games. Q. J. Econ. 107 (2), 739–771.
Cooper, D.J., Kagel, J.H., 2008. Learning and transfer in signaling games. Econ. Theory 34 (3), 415–439.
Cooper, D.J., Kagel, J.H., 2009. The role of context and team play in cross-game learning. J. Eur. Econ. Assoc. 7 (5), 1101–1139.
Crawford, V.P., 1995. Adaptive dynamics in coordination games. Econometrica 63 (1), 103–143.
Dal Bó, P., Fréchette, G.R., 2011. The evolution of cooperation in infinitely repeated games: experimental evidence. Am. Econ. Rev. 101 (1), 411–429.
Dal Bó, P., Fréchette, G.R., 2018. On the determinants of cooperation in infinitely repeated games: a survey. J. Econ. Lit. 56 (1), 60–114.
Dubois, D., Willinger, M., Van Nguyen, P., 2012. Optimization incentive and relative riskiness in experimental stag-hunt games. Int. J. Game Theory 41 (2), 369–380.

Duffy, J., Feltovich, N., 2002. Do actions speak louder than words? An experimental comparison of observation and cheap talk. Games Econ. Behav. 39 (1), 1–27.
Embrey, M., Fréchette, G.R., Yuksel, S., 2017. Cooperation in the finitely repeated prisoner's dilemma. Q. J. Econ. 133 (1), 509–551.
Feltovich, N., Iwasaki, A., Oda, S.H., 2012. Payoff levels, loss avoidance, and equilibrium selection in games with multiple equilibria: an experimental study. Econ. Inq. 50 (4), 932–952.
Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.
Fréchette, G.R., 2009. Learning in a multilateral bargaining experiments. J. Econom. 153 (2), 183–195.
Fréchette, G.R., 2012. Session-effects in the laboratory. Exp. Econ. 15 (3), 485–498.
Fréchette, G.R., Lizzeri, A., Perego, J., 2020. Rules and commitment in communication: an experimental analysis. Working Paper.
Gueth, W., Huck, S., Rapoport, A., 1998. The limitations of the positional order effect: can it support silent threats and non-equilibrium behavior? J. Econ. Behav. Organ. 34, 313–325.
Harrison, G., Martínez-Correa, J., Swarthout, T., 2013. Inducing risk neutral preferences with binary lotteries: a reconsideration. J. Econ. Behav. Organ. 94, 145–159.
Harsanyi, J.C., Selten, R., 1988. A General Theory of Equilibrium Selection in Games. MIT Press, Cambridge, MA.
Healy, P.J., 2016. Epistemic experiments: utilities, beliefs, and irrational play. Working Paper.
Heinemann, F., Nagel, R., Ockenfels, P., 2009. Measuring strategic uncertainty in coordination games. Rev. Econ. Stud. 76 (1), 181–221.
Hoffman, E., McCabe, K.A., Smith, V., 1998. Behavioral foundations of reciprocity: experimental economics and evolutionary psychology. Econ. Inq. 36, 335–352.
Huck, S., Jehiel, P., Rutter, T., 2011. Feedback spillover and analogy-based expectations: a multi-game experiment. Games Econ. Behav. 71 (2), 351–365.
Kandori, M., Mailath, G.J., Rob, R., 1993. Learning, mutation, and long run equilibria in games. Econometrica 61 (1), 29–56.
Kartal, M., Müller, W., 2018. A new approach to the analysis of cooperation under the shadow of the future: theory and experimental evidence. Working Paper.
Kendall, R., 2020. Decomposing coordination failure in stag hunt games. Working Paper.
Kirchkamp, O., Oechssler, J., Sofianos, A., 2021. The binary lottery procedure does not induce risk neutrality in the Holt & Laury and Eckel & Grossman tasks. J. Econ. Behav. Organ. 185, 348–369.
Levine, D.K., Palfrey, T.R., 2007. The paradox of voter participation? A laboratory study. Am. Polit. Sci. Rev. 101 (1), 143–158.
Luce, R.D., Raiffa, H., 1957. Games and Decisions: Introduction and Critical Survey. John Wiley & Sons, New York, NY.
Proto, E., Rustichini, A., Sofianos, A., 2021. Intelligence, errors and strategic choices in the repeated prisoners' dilemma. Working Paper.
Rankin, F.W., Van Huyck, J.B., Battalio, R.C., 2000. Strategic similarity and emergent conventions: evidence from similar stag hunt games. Games Econ. Behav. 32 (2), 315–337.
Rick, S., Weber, R.A., 2010. Meaningful learning and transfer of learning in games played repeatedly without feedback. Games Econ. Behav. 68 (2), 716–730.
Roth, A.E., Malouf, M.W., 1979. Game-theoretic models and the role of information in bargaining. Psychol. Rev. 86 (6), 574–594.
Schelling, T.C., 1960. The Strategy of Conflict. Harvard University Press, Cambridge, MA.
Schmidt, D., Shupp, R., Walker, J.M., Ostrom, E., 2003. Playing safe in coordination games:: the roles of risk dominance, payoff dominance, and history of play. Games Econ. Behav. 42 (2), 281–299.
Selten, R., 1995. An axiomatic theory of a risk dominance measure for bipolar games with linear incentives. Games Econ. Behav. 8 (1), 213–263.
Straub, P.G., 1995. Risk dominance and coordination failures in static games. Q. Rev. Econ. Finance 35 (4), 339–363.
Van Huyck, J.B., Battalio, R.C., 2002. Prudence, justice, benevolence, and sex: evidence from similar bargaining games. J. Econ. Theory 104, 227–246.
Van Huyck, J.B., Battalio, R.C., Beil, R.O., 1990. Tacit coordination games, strategic uncertainty, and coordination failure. Am. Econ. Rev. 80 (1), 234–248.
Vespa, E., Wilson, A.J., 2017. Experimenting with equilibrium selection in dynamic games. Working Paper.
Young, H.P., 1993. The evolution of conventions. Econometrica 61 (1), 57–84.