Journal of Behavioral and Experimental Economics 113 (2024) 102290

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Coordination in stag hunt games

Rui Silva

School of Economics, University of East Anglia, United Kingdom

ARTICLE INFO

JEL classification:
C72
D81

Keywords:
Equilibrium selection
Coordination
Relative salience
Stag hunt game
Risk dominance
Payoff dominance

ABSTRACT

Stag hunt games display a tension between a payoff superior option (stag) and a less risky but payoff inferior alternative (hare). We explore that tension by proposing a selection criterion (which we denote as *relative salience*) where subjects choose to coordinate in one option by comparing the salience of stag's main aspect (its payoff) to the salience of hare's main aspect (its risk) by accounting for both payoff-relevant factors and unobservable individual-specific (idiosyncratic) preferences. Using data from 10 experiments, we find that this criterion is a significant determinant of individual choices in stag-hunt games, outperforming other selection methods.

1. Introduction

Stag hunt games (SHG) are the prototype of a social contract, capturing the main strategic forces and elements present in several economic problems, such as currency attacks, bank runs, asset bubbles, or technology diffusion. In their simplest form, as depicted in Fig. 1, they are played by two subjects with two options (stag and hare). That game features two Pareto-ranked equilibria in pure actions (i.e., when each player picks one option with 100% certainty). One where each subject picks stag; and another where, instead, they select hare. The former is said to payoff dominate the latter since $a > b$.

	stag	hare
stag	a, a	c, d
hare	d, c	b, b

Fig. 1. Stag hunt game where $d < a > b > c$.

SHGs display a tension between stag (the most efficient option) and hare (the less "risky" option). An option's riskiness (or the robustness to strategic uncertainty), say hare, depends on the maximum probability with which the opponent may choose stag, and hare still be the best response to it. That probability is referred to as the size of the basin of attraction. When it is larger than 1/2, we say that hare is risk-dominant.

That tension has been reported in different experiments. In some cases, subjects seem to favor efficiency (stag) over risk (hare) while, in others, it is the other way around. Often, individuals seem to be initially more likely to attempt coordination in the most efficient equilibrium (i.e., the payoff dominant outcome (stag,stag)) while frequently converging to the risk-dominant equilibrium as a game is repeated. Examples can be found in Battalio et al. (2001), Clark et al. (2001), Dal Bó et al. (2021), Dubois et al. (2012), Schmidt et al. (2003), and Straub (1995).

A fair share of attention has been devoted to the reasons for coordination failure in most games, i.e., the tendency of subjects to converge to the Pareto inferior option (hare), but not so much why, in some games, they pick stag, and why that option's frequency varies so much. We argue that the observed behavior in different experiments can be explained as the product of the referred tension or trade-off between risk and efficiency. We propose a selection criterion based on that hypothesis which depends on payoff-relevant factors (i.e., those in the payoff matrix) and unobservable individual-specific preferences (unobserved heterogeneity).

To the best of my knowledge, the only other equilibrium selection methods for 2 × 2 coordination games that assume that individuals solely rely on reasoning, deduction, and focal aspects considering the information in the payoff matrix (thus, also often denoted as *deductive methods*) are risk dominance (r) (Harsanyi & Selten, 1988), the optimization premium (op) (Battalio et al., 2001), and relative riskiness (rr) (Dubois et al., 2012). Those methods are often used to justify coordination failure (i.e., the failure to coordinate on stag) when games are repeated. However, they are frequently unable to explain why subjects, at times, coordinate on stag, and why that option's frequency varies so much across games.

For instance, in the games in Fig. 2, the r, op, and rr are all the same, indicating an equivalent level of strategic uncertainty across those

E-mail address: r.constantino-da-silva@uea.ac.uk.

https://doi.org/10.1016/j.socec.2024.102290
Received 1 August 2023; Received in revised form 6 August 2024; Accepted 11 September 2024
Available online 14 September 2024

Game 1

	stag	hare
stag	62, 62	48, 51
hare	51, 48	61, 61

Game 2

	stag	hare
stag	97, 97	13, 86
hare	86, 13	26, 26

Fig. 2. Games 1 and 2 from Kendall (2022).

games. We return to these games in Section 2 and explain how those values were calculated. At this stage, however, it is enough to highlight that the frequency of stag in each game is very different: 64% in game 1 and 81% in game 2 in round 1; 50% and 88% respectively by round 75. Since each option's risk is, in theory, the same, the higher frequency of stag in game 2 seems to reflect larger differences between each equilibrium's payoff. This example illustrates the apparent existence of a trade-off between risk (choosing hare) and efficiency (choosing stag) by highlighting how larger differences between each equilibrium's payoffs seem to drive subjects to choose stag when hare's risk is the same. We hypothesize that a similar dynamics holds if, instead, the differences between payoffs are fixed and only the risk would vary.

A similar intuition is present in Kendall (2022).[1] They decompose SHGs into different complementary substructures (or components) to identify the source of the limitations of existent criteria in explaining observed behavior and to develop, in the process, a toolbox that would enable the construction of equivalence classes of games (in essence, fundamentally framing in a different way the decomposition proposed in Candogan et al. (2011) as we discuss in Section 2). They find that the information in one of those components seems to be missing from the above criteria. However, further reading into Kendall (2022) leads us to believe that the difference between each equilibrium's payoff (encompassed in that component) is the main aspect not captured by those criteria. Considering the richer experimental set in our paper, we find that the information in that component is not a statistically significant predictor of subjects' behavior, unlike the difference between the equilibrium payoffs.

Considering the intuition above, in the first part of the paper, we propose a selection criterion (*relative salience*) where choices in a SHG reflect a tension between payoff-relevant factors and depend on the salience of each option's key attribute (stag's payoff or hare's risk). The salience of an option represents the probability that a subject selected at random from a population would select that alternative considering its defining aspect and individual-specific preferences and tastes (which are not observed by the experimenter). The defining aspect of stag is its payoff while risk is the defining aspect of hare. The more salient one option's defining aspect, the more focal it is and the more likely are subjects to pick that alternative by the end of an experiment.

In the second part of the paper, we consider data from 10 experiments in SHG covering 33 different game structures, 182 experimental sessions, and 2528 individual choices. We find that our criterion accounts for the observed behavior at the individual and session levels, outperforming other selection methods. We observe, nonetheless, that this criterion's relevance and predictive power crucially depend on how salient one option is relative to another. In particular, convergence to the stag equilibrium appears to require larger levels of salience relative to hare compared to the salience needed to ensure convergence to the hare equilibrium, an aspect consistent with the tendency of most subjects to converge to the risk-dominant option by the end of an experiment.

In the remainder of the paper, we proceed as follows: in Section 2, we describe existing equilibrium selection and game decomposition methods. In Section 3, we introduce our criterion and outline our research hypotheses. In Section 4, we describe the experimental data and methods used to test the hypotheses in Section 2, and in Section 5, we interpret the respective results. In Section 6, we summarize the main findings and provide some concluding remarks.

2. Equilibrium selection and game decomposition

Using the game in Fig. 1 as a reference, based on Harsanyi and Selten (1988), we say that (stag, stag) is payoff dominant (or simply that stag is payoff dominant) if $a > b$. We say that (hare, hare) is risk dominant (or just that hare is risk dominant) if $(b - c) > (a - d)$. Equivalently, hare is risk dominant when its basin of attraction $r = (b - c)/(a - d + b - c)$ is larger than $1/2$.[2] As r increases, so does the maximum probability with which the opponent may choose stag against which hare is still a best response. The larger the r, the more robust is hare to the strategic uncertainty underlying the game, and thus, the more risk-dominant that alternative is.

The optimization premium (Battalio et al., 2001) is another criterion that attempts to capture a game's strategic uncertainty. In it, we assume that subjects are more likely to choose HARE when the incentive to best respond to the opponent increases (i.e., in their words, when the optimization premium is high), and thus, when a subject obtains larger payoff losses from not choosing stag when the opponent does so, or hare when that is the option selected by the opponent. Using the game in Fig. 1, we define the optimization premium as $op = (a - d) + (b - c)$.

Relative riskiness (Dubois et al., 2012) is a more recent criterion that mainly focuses on the security aspect of play. It expresses how much a subject's payoff varies, after picking hare, if the opponent does not best respond to it, compared to the respective loss if the subject would have picked stag and the opponent would not best respond to it. We define relative riskiness as $rr = |b - d|/(a - c)$.

In the games depicted in Fig. 2, we obtain that $r \approx 0.542$, $op = 24$, and $rr \approx .714$. The differences in the frequency of stag in each game highlight the limitations of those methods to account for the observed behavior in SHGs. In a recent paper, Kendall (2022) attempts to identify the source of such limitations by decomposing SHGs into three sub-structures: the kernel, strategic, and behavioral components.

Using Fig. 1 as a reference, the kernel represents a subject's average payoff over every possible outcome. An outcome's kernel value can be written as $k = (a + b + c + d)/4$. In symmetric games, it represents an additive scalar factor to a subject's payoff function. However, the information in the kernel is not useful for a subject from a strategic perspective.

The strategic component captures the effect of payoff-relevant information on a subject's choice. It has value from a strategic angle since it mainly captures the intuition underlying the idea of risk dominance. We define the strategic values of stag and hare as $s_1 = a - (a + d)/2$ and $s_2 = b - (b + c)/2$, respectively.

[1] An approach with a similar intuition, using instead Prisoner's Dilemma (PD) games, is found in Mengel (2018). That paper attempts to disentangle the role of risk (the penalty for deviating from the unique equilibrium in that game) and temptation (the gain for deviating from the Pareto superior outcome) while also accounting for the efficiency (or the payoff) from cooperating with an opponent. However, rather than a trade-off, Mengel (2018) tests which aspect better explains the observed behavior in one-shot and repeated PD games.

[2] The size of the basin of attraction is proportional to the probability under which stag is played in a mixed equilibrium (i.e., a Nash equilibrium where each player chooses stag with a probability lower than 100%).

	stag	hare		kernel		+	strategic		+	behavioral	
stag	a, a	c, d	=	k, k	k, k	+	s_1, s_1	$-s_2, -s_1$	+	b, b	$-b, b$
hare	d, c	b, b		k, k	k, k		$-s_1, -s_2$	s_2, s_2		$b, -b$	$-b, -b$

Fig. 3. Decomposition of a SHG into kernel ($k = (a+b+c+d)/2$), strategic ($s_1 = a-(a+d)/2$ and $s_2 = b-(b+c)/2$), and behavioral ($b = [(a+d)-(b+c)]/4$) components in Kendall (2022).

	stag	hare		strategic		+	non-strategic	
stag	a, a	c, d	=	$a-nc_1, a-nc_1$	$c-nc_2, d-nc_1$	+	nc_1, nc_1	nc_2, nc_1
hare	d, c	b, b		$d-nc_1, c-nc_2$	$b-nc_2, b-nc_2$		nc_1, nc_2	nc_2, nc_2

Fig. 4. Decomposition of a SHG into strategic and non-strategic ($nc_1 = (a+d)/2$ and $nc_2 = (c+b)/2$) components in Candogan et al. (2011).

Finally, the behavioral component represents the difference between the average game payoff (the kernel) and the average payoff when the opponent commits to an action. We define the behavioral value as $b = (a+d)/2-(a+b+c+d)/4$ or just $[(a+d)-(b+c)]/4$. As with the kernel, the information in b does not have value from a strategic perspective, and its interpretation, as a consequence, is extremely challenging. Kendall (2022) finds that b is missing in each of the selection criteria above and that such is the reason for their limitations in accounting for observed behavior in SHGs. We believe, however, that the information not captured by those criteria, and partially reflected in the behavioral component, may simply be the difference between each equilibrium's payoff, an aspect carrying strategic value since it can be used as a coordination device. That conjecture finds support in the fact that b does not emerge as a statistically significant determinant of stag choices when we consider a richer experimental set like the one in this paper, unlike the difference between each equilibrium's payoff (see both the results in Section 4 and Appendix 2). The decomposition of a SHG as proposed in Kendall (2022) is depicted in Fig. 3.

The decomposition in Kendall (2022) is closely related to the direct-sum decomposition of finite games into strategic and non-strategic components in Candogan et al. (2011). The non-strategic component can be interpreted as the average payoff when a player picks one option at random fixing the opponent's choice. Considering the game in Fig. 1, when we fix the column's player at stag and hare, row's payoffs are $nc_1 = (a+d)/2$ and $nc_2 = (c+b)/2$ in the non-strategic component. The strategic component is obtained by subtracting the non-strategic component from each payoff in the original game. We depict those two components in Fig. 4.

We can see that in a 2 × 2 coordination game, the non-strategic component in Candogan et al. (2011) is simply the sum of the kernel and behavioral components in Kendall (2022) while their strategic components coincide. Therefore, it suffices to test one of the approaches as a possible determinant of stag choices (we opted for Kendall (2022) given the finer level of disaggregation of their components). In a recent working paper, Garcia-Galocha et al. (2024) test, among other things, whether the non-strategic component in Candogan et al. (2011) helps to explain the observed behavior (and thus, deviations from equilibrium play) in 3 × 3 variations of a prisoner's dilemma games. They propose a solution concept (*Mutual-Max Sum or simply MMS)* to identify the expected outcome of a game when assuming that strategic and non-strategic components condition behavior. In the MMS, subjects choose actions to maximize the sum of the other player's payoff, thus expressing an extreme form of altruism and empathy. In the game in Fig. 1, subjects would coordinate in stag if $m_1 = a+d > b+c = m_2$, and on hare otherwise, where m_1 and m_2 denote the MMS value of each action. Garcia-Galocha et al. (2024) find that the MMS prediction is only relevant when it coincides with the equilibrium prediction in the game in Fig. 4 with non-strategic components (and thus when empathy and altruism reinforce non-strategic aspects of the game in a possible reading of that result). Unfortunately, in the game in Fig. 1, the MMS prediction will always coincide with one equilibrium in the game with non-strategic components (meaning that we cannot test their prediction). Moreover, in Section 5, we find that the non-strategic components in Kendall (2022) are not relevant in explaining the observed behavior in SHGs. Hence, we would not expect the MMS concept to be it either. In Table 10 [Appendix 2], we show that such is precisely the case.

3. Relative salience

Stag hunt games display a tension between efficiency (stag) - choosing an action with the highest payoff in equilibrium - and risk (hare) - minimizing the payoff loss from not best responding to the opponent. We propose an equilibrium selection criterion (denoted as *relative salience*) where choices reflect a tension between those two aspects and depend on the salience of each option (stag or hare). By salience, as more formally defined below, we refer to the probability that a subject randomly selected from a population, selects one option based on its distinctive feature (i.e., the payoff in the case of stag, and the risk in the case of hare).

In what follows, we normalize the payoffs of the game in Fig. 1 in the unit interval through a linear transformation, such that $a = 1$ and $c = 0$, and thus, $\alpha = (b-c)/(a-c)$ and $\beta = (d-c)/(a-c)$. The normalized payoff matrix is depicted in Fig. 5.

	stag	hare
stag	$1, 1$	$0, \beta$
hare	$\beta, 0$	α, α

Fig. 5. Normalized stag hunt game.

Let $\gamma = \{\text{stag}, \text{hare}\}$ be an action and $\delta = \{\text{PAYOFF}, \text{RISK}\}$ the feature used to evaluate γ. We define the RISK of γ as in Harsanyi and Selten (1988), and thus, as the payoff loss following a unilateral deviation. When $\gamma = \text{hare}$, the loss is α. When $\gamma = \text{stag}$, it amounts to $1-\beta$.

Feature δ is observable. However, in most SHGs (and virtually, most experiments), it is sensible to assume that there are features or attributes of an option (either stag or hare) that are unobservable to the experimenter but that shape a subject's preferences (i.e., are subject-specific). For example, different subjects may have distinct purposes for the monetary payment from participating in the experiment (e.g., they may be looking to buy a pair of sneakers, go out with their friends, or save money for the summer break). Such observation introduces unobserved heterogeneity across the participants' preferences, thus making the risky option stag more or less attractive for each specific subject.

In that sense, we define the attractiveness or value of γ for subject i conditional on δ as an additive combination of observable (payoff-specific) features and unobservable (subject-specific) preferences as in a random utility model as

$$s_i(\gamma|\delta) = x_\gamma^\delta + \varepsilon_{i,\gamma} \tag{1}$$

where $\varepsilon_{i,\gamma}$ is a random variable (or noise) representing *unobservable attributes* or idiosyncratic preferences over γ for subject i while x_γ^δ represents its *observable attributes* conditional on feature δ. Therefore:

- if δ = PAYOFF then $x_{\text{stag}}^\delta = 1$ and $x_{\text{hare}}^\delta = \alpha$;
- if δ = RISK then $x_{\text{stag}}^\delta = 1 - \beta$ and $x_{\text{hare}}^\delta = \alpha$.

It follows that the value of hare is always $\alpha + \varepsilon_i$ independently of the feature being the PAYOFF or the RISK. The value of stag is $1 + \varepsilon_i$ if we assess that option considering its PAYOFF, and decreases by β if, instead, we look at its RISK. It means that the attractiveness of hare only depends on its RISK (given by α, which is simultaneously its PAYOFF) independently of the feature used to assess it, varying depending on individual unobservable tastes and preferences. The attractiveness of stag varies depending on the feature being the PAYOFF or the RISK. We say that subject i prefers γ over γ' conditional on δ if

$$\varepsilon_{i,\gamma'} < \left(x_\gamma^\delta - x_{\gamma'}^\delta\right) + \varepsilon_{i,\gamma} \tag{2}$$

The probability of γ being chosen by a subject selected at random from a population can be defined as

$$\Pr(\gamma|\delta) = \Pr\left(\varepsilon_{i,\gamma'} < \left(x_\gamma^\delta - x_{\gamma'}^\delta\right) + \varepsilon_{i,\gamma}\right) \tag{3}$$

which derives from the probability of γ being chosen conditional on feature δ and $\varepsilon_{i,\gamma'}$ by integration of the latter when accounting for its marginal density $f_{\varepsilon_{i,\gamma'}}$. The respective conditional probability can be written as

$$\Pr(\gamma|\delta, \varepsilon_{i,\gamma'}) = F\left(\left[x_\gamma^\delta - x_{\gamma'}^\delta\right] + \varepsilon_{i,\gamma}\right) \tag{4}$$

where F denotes the cumulative distribution function of subject-specific disturbances $\varepsilon_{i,\gamma'}$ for every i.

We assume that each subject picks an option considering the salience of its distinctive feature. By distinctive feature, we mean the PAYOFF if γ = stag, and RISK if γ = hare. By salience, we refer to how likely an individual randomly selected from a population is to pick γ relative to its alternative γ' considering the distinctive feature of the former (i.e., γ).

In the same vein as a discrete choice model, we assume that individual (unobservable) disturbances are independent and identically distributed across subjects (which not only implies the homoscedasticity of those error terms but also seems a reasonable assumption considering how participants are selected for an experiment and how it is run) according to a Gumbel distribution. In that case, we can define the probability of picking either stag or hare and thus, the salience of each option, as closed-forms logit transformations of the deterministic parts of s_i(stag|PAYOFF) and s_i(hare|RISK), such that

$$\sigma(\text{stag}|\text{PAYOFF}) = \frac{e^{(1)}}{e^{(1)} + e^{(\alpha)}} \quad \text{and} \quad \sigma(\text{hare}|\text{RISK}) = \frac{e^{(\alpha)}}{e^{(\alpha)} + e^{(1-\beta)}} \tag{5}$$

Conditional on individual-specific tastes and disturbances, we assume that subject i picks the option that provides the best trade-off between RISK and PAYOFF. For example, if that option is stag, then it maximizes the salience of its distinctive feature (the excess payoff compared to hare) and, as its dual, minimizes the salience deficit of its non-distinctive aspect (the excess risk taken compared to hare). We represent that trade-off as

$$\sigma = \frac{\sigma(\text{stag}|\text{PAYOFF})}{\sigma(\text{hare}|\text{RISK})} \tag{6}$$

We denote σ as the *relative salience* of a SHG. When $\sigma > 1$, subjects are more likely to choose stag; otherwise, they are more likely to pick hare. In Appendix 2, we compare σ to other functional specifications (e.g., by summing or multiplying σ(stag|PAYOFF) and σ(hare|RISK)) but do not find any that outperforms it.

In the next section, we check how well σ accounts for the observed behavior in different experiments. Additionally, we test the importance of unobserved attributes by comparing σ to a version where we omit the stochastic element from $s_i(\gamma|\delta)$, and thus, ignore the role of subject-specific (unobservable) tastes and disturbances in the likelihood of choosing option γ. The resulting ratio mimics the matching law in Luce's choice axiom, which we denote as *Luce's relative salience* and write as

$$\sigma^l = \frac{1/(1+\alpha)}{\alpha/(\alpha+\beta-1)} \tag{7}$$

Considering this framework, we outline several hypotheses that we test using a large experimental set. Ignoring other factors that could affect beliefs and consequently, subjects' choices, our first hypothesis is that subject i is more likely to choose the most salient option.

Hypothesis 1. The larger is σ, the greater the probability of stag being chosen.

As subjects play repeatedly against different opponents, we expect their beliefs to be gradually adjusted around the focal aspects of a game. In that case, we conjecture that σ acts as a coordination device, especially accounting for choices at later stages of a game. Despite the noise underlying beliefs in the initial rounds of a game, we still predict σ to be an important, though weaker, predictor of behavior.

Hypothesis 2. As a game is repeated, the marginal effect of changes in σ in the probability of choosing stag increases. Despite the lower marginal effect in the first round, σ is a statistically significant determinant of behavior.

When one option's salience is close to another, i.e., when σ is close to 1, other aspects may shape a subject's beliefs about the opponent, and affect that subject's choice. In those cases, we expect σ to be less correlated with the observed behavior, and we say that stag or hare is *weakly salient*; otherwise, we say that they are *strongly salient*. Considering the characteristics of our experimental data, we test four different thresholds τ separating strong from weak salience: 5%, 10%, 15%, and 20%. In each case, we say that stag (hare) is strongly (weakly) salient if its PAYOFF (RISK) is τ more (less) salient than hare's (stag's) RISK (PAYOFF).

Hypothesis 3. The weaker the relative salience of stag, the less likely is that σ is a statistically significant determinant of that option's frequency.

Considering that in most experiments, games frequently converge to the risk-dominant equilibrium (hare), it is sensible to conjecture that convergence to each equilibrium requires distinct minimum levels of salience. In that case, we hypothesize that the convergence to stag requires a larger τ than hare.

Hypothesis 4. Convergence to stag requires a larger τ than convergence to hare.

Table 1 summarizes the main parameters of each game in Fig. 2. In game 2, we would predict subjects to pick stag (which is strongly salient even with a 20% threshold). That option was, in fact, chosen by the majority of subjects. In game 1, despite $\sigma < 1$, hare is only weakly salient (even at a threshold of 5%). Therefore, we would not expect that option to be more frequently chosen than stag. That is, exactly, what we observe, given that only 50% of the subjects choose hare by round 75. Moreover, in each of the four sessions using that game, the observed choices point against any behavioral consistency. In one session, almost every subject chose hare. In another session, everyone picked stag by round 75. In the remaining two sessions, choices were almost evenly split.

Table 1
Parameters and salience in the games from Fig. 2.

Parameters	Game 1	Game 2
α	.928	.154
β	.214	.869
σ	.966	1.382
Most salient option	hare (weakly when $\tau = 5\%$)	stag (strongly when $\tau = 20\%$)

4. Data & empirical strategy

We collected data from ten published papers with experiments satisfying the following conditions:

(a) use a (two-player) 2×2 symmetric stag-hunt games in which stag is payoff dominant and hare is risk dominant.
(b) each experimental session does not involve pre-play communication and employs a non-fixed (exogenous) rematching protocol between participants.

Table 2
Main elements of each experiment.

Paper	Games	id	α	β	σ	Sessions	Periods	Players
Cooper et al. (1992)	1					3	22	33
		1	.8	.8	.851			
Straub (1995)	4					4	9	40
		2	.8	.8	.851			
		3	.5	.5	1.244			
		4	.75	.75	.903			
		5	.285	.857	1.254			
Battalio et al. (2001)	3					24	75	192
		6	.266	.933	1.229			
		7	.444	.888	1.091			
		8	.888	.777	.798			
Clark et al. (2001)	3					5	10	100
		1	.8	.8	.851			
		9	.7	.9	.889			
		10	.9	.7	.812			
Duffy and Feltovich (2002)	1					3	10	60
		11	.75	.75	.903			
Schmidt et al. (2003)	4					16	8	160
		12	.5	.5	1.244			
		13	.75	.75	.903			
		14	.5	.5	1.244			
		15	.6	.8	1			
Dubois et al. (2012)	3					24	75	192
		6	.888	.777	1.229			
		16	.6	.85	.980			
		17	.6	.85	.980			
Feltovich et al. (2012)	6					30	20–40	486
		18	.667	.667	1			
		19	.667	.667	1			
		20	.833	.833	.819			
		21	.687	.687	.974			
		22	.687	.687	.974			
		23	.687	.687	.974			
Dal Bó et al. (2021)	8					65	15	1140
		24	.625	.437	1.149			
		25	.625	.687	1.025			
		26	.571	.642	1.094			
		27	.5	.583	1.196			
		28	.625	.937	.929			
		29	.571	.928	.972			
		30	.5	.916	1.033			
		31	.4	.9	1.123			
Kendall (2022)	2					8	75	128
		32	.928	.214	.966			
		33	.154	.869	1.384			
Total						182	6265	2531

To the best of my knowledge, the papers in Table 2 (which also describes the characteristics of each experiment) are the only published articles conducting experiments with games and sessions satisfying

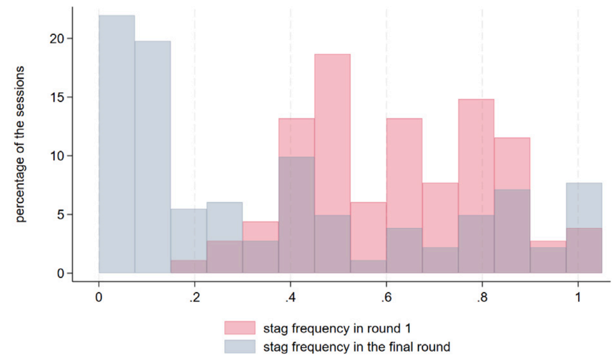

Fig. 6. Frequency of stag choices at the session level.

conditions (a) and (b). Our sample contains 33 different games, 182 sessions, and 2531 subjects who played those games repeatedly between 9 and 75 times. In 48% of the sessions, stag's payoff is more salient than hare's risk ($\sigma > 1$). Relative salience varies substantially, ranging from hare being 20% more salient than stag to the latter being 39% more salient than the former. In 58% of the sessions, one option is, at least, 5% more salient than the other. That figure falls down to 41% when $\tau = 10\%$, and 27% when $\tau = 20\%$. The games and main parameters are available in Appendix 1.

Table 3
Summary statistics (frequency of stag) of each experiment.

Paper	Sessions	R1	RQ1	RQ2	RQ3	RF
Cooper et al. (1992)	3					
mean		.566	.266	.066	0	.033
std dev		.057	.208	.057	0	.057
Straub (1995)	4					
mean		.525	.5	.55	.475	.475
std dev		.298	.294	.479	.499	.55
Battalio et al. (2001)	24					
mean		.625	.348	.307	.270	.244
std dev		.132	.333	.366	.377	.332
Clark et al. (2001)	5					
mean		.43	.24	.19	.02	.1
std dev		.152	.065	.108	.044	.070
Duffy and Feltovich (2002)	3					
mean		.683	.666	.666	.6	.483
std dev		.115	.202	.208	.217	.202
Schmidt et al. (2003)	16					
mean		.612	.687	.662	.693	.656
std dev		.192	.212	.189	.306	.361
Dubois et al. (2012)	24					
mean		.765	.510	.401	.375	.390
std dev		.166	.303	.364	.420	.346
Feltovich et al. (2012)	30					
mean		.795	.717	.648	.623	.563
std dev		.148	.222	.292	.317	.311
Dal Bó et al. (2021)	65					
mean		.477	.383	.327	.290	.240
std dev		.131	.185	.213	.223	.228
Kendall (2022)	8					
mean		.726	.703	.703	.734	.687
std dev		.194	.353	.357	.356	.329
Total	182					
mean		.614	.492	.435	.405	.372
std dev		.197	.284	.321	.355	.338

The main benefit of combining data from different experiments is the possibility of using a reasonable variety of payoff structures and groups of subjects. We recognize, nonetheless, that differences in the design and running of each experiment could have possibly affected the observed behavior and even confounded the results. In any case, in our analysis, we attempt to minimize such effects by controlling for each experiment (article) specific factors.

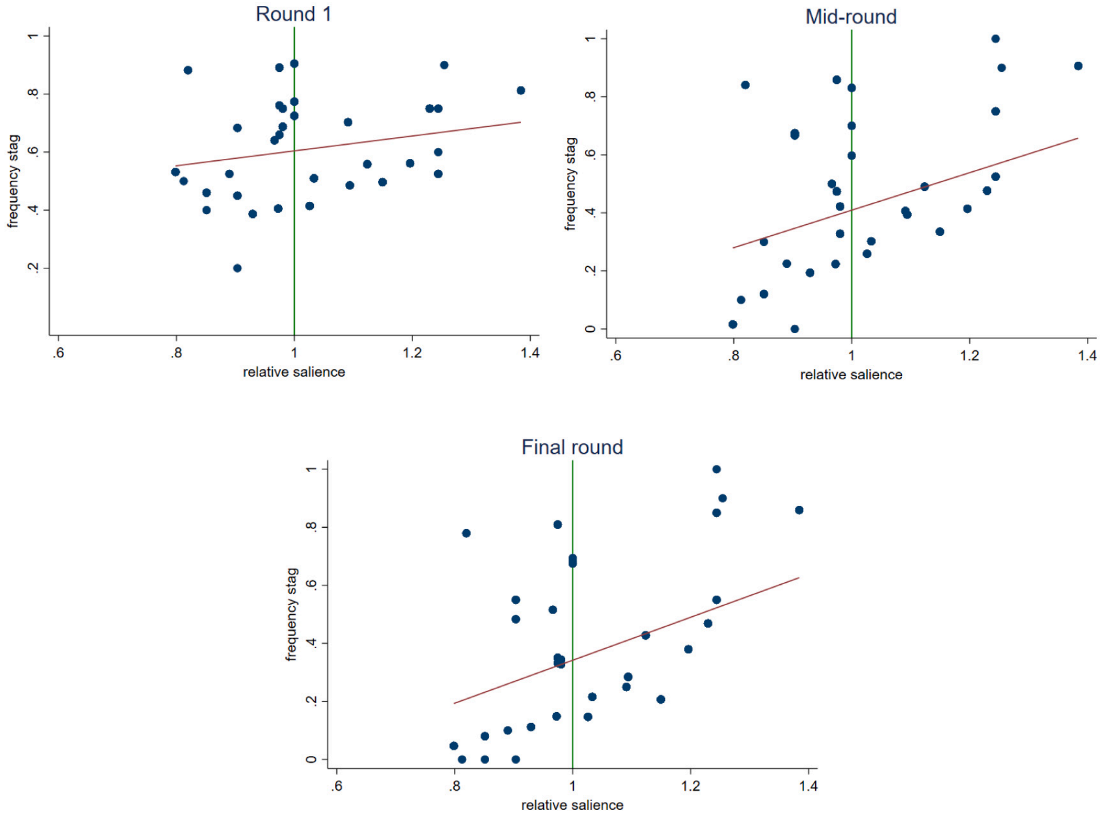

Fig. 7. Average frequency of stag in each game (represented by a dot) against relative salience in round one, mid-round, and the final round. In red, we include a linear line of best fit.

In our sample, at the session level, we confirm the propensity for efficient play in the first round (Fig. 6). In 77% of the sessions, at least 50% of the subjects chose stag. As the games were repeated, play converged to the risk-dominant option. In the final round, 71% of the sessions have not less than half of the participants choosing hare. Despite that scenario, we note that in 18% of the sessions, stag is still played by not less than 80% of the subjects in the final round of a game.

Table 3 presents the average share and dispersion of stag choices in each article at different stages of an experiment. We do not find it necessarily sensible to compare sessions from distinct experiments at specific rounds (e.g., Dal Bó et al. (2021) compares choices in rounds 1 and 8 in some of the experiments in our sample) because subjects are often told at the beginning of a session how many periods will be played. For that reason, instead, we consider different points in each experiment: the first period (R1); after 25% of the periods had been played (RQ1); after 50% of the periods had been played (RQ2); after 75% of the periods had been played (RQ3); and the final period (RF). We use those partitions when testing the correlation between σ and stag choices.

The plots in Fig. 7 show a correlation between σ and the frequency of stag (as proposed in Hypothesis 1) which appears to be more pronounced in the later rounds of a game (consistent with Hypothesis 3). The correlation between other selection criteria and stag choices (Figure 9 [Appendix 2]) seems weaker and the scatter is more dispersed. The dispersion in an experiment's final round in Fig. 7 also lends support to Hypothesis 4, and thus, that convergence to stag requires a larger salience than the convergence to hare. In particular, in games where stag was selected by at least 50% of subjects, we find that $\sigma > 1.2$ while, with some exceptions (four games), most subjects choose hare as long as $\sigma < 1$.

In that sense, we further inspect the prevalence of stag choices at different levels of relative salience σ. As depicted in Fig. 8, on average, we require a $\sigma > 1.15$ to guarantee that the frequency of stag remains above 0.5. On the other hand, most subjects do not seem to require too many rounds to pick hare as long as $\sigma < 1$.

We test the hypotheses in Section 3 using a probit model at the subject level

$$Y^t_{i,s,a} = \beta_1 + \beta_2\sigma_{s,a} + \rho X + \lambda r_{s,a} + \alpha_a + u^t_{i,s,a} \quad (8)$$

where $Y^t_{i,s,a} = 1$ if subject i picked stag in session s of the article a in round $t = \{R1, RQ1, RQ2, RQ3, RF\}$, $\sigma_{s,a}$ is the relative salience of a game in session s from an article a, $X = \{r_{s,a}, op_{s,a}, rr_{s,a}, b_{s,a}, s_{1,s,a}, s_{2,s,a}, \sigma^l_{s,a}\}$ is a matrix of competing selection criteria (respectively, basin of attraction, optimization premium, relative riskiness, behavioral component, strategic component of stag, strategic component of hare, and Luce's relative riskiness), $r_{s,a}$ indicates the number of periods in a session s on an article a, and finally, α_a controls for article-specific factors, i.e., article fixed-effects (which simultaneously controls for differences in the number of players, region, and type of participant).

5. Results

The correlation between σ and individual choices at different points of a session (Table 4) is consistent with Hypotheses 1 and 2, and thus, indicates that σ is a significant determinant of the prevalence of stag. At the same time, the marginal impact of changes in σ in the probability of picking stag increases as a game is repeated, which favors the conjecture underlying Hypothesis 4.

In round 1, on average, an increase of 1 p.p. in the salience of stag's payoff relative to hare's risk increases the probability of a subject picking the former option by approximately 0.44 p.p. By the final round, that effect increases to 0.82 p.p. The difference between those two marginal effects is significant at 1%. The latter two columns in Table 4 highlight the importance of subject-specific tastes and disturbances as Luce's relative salience σ^l is not a statistically significant predictor of stag choices compared to σ. In Appendix 2, we compare σ to other

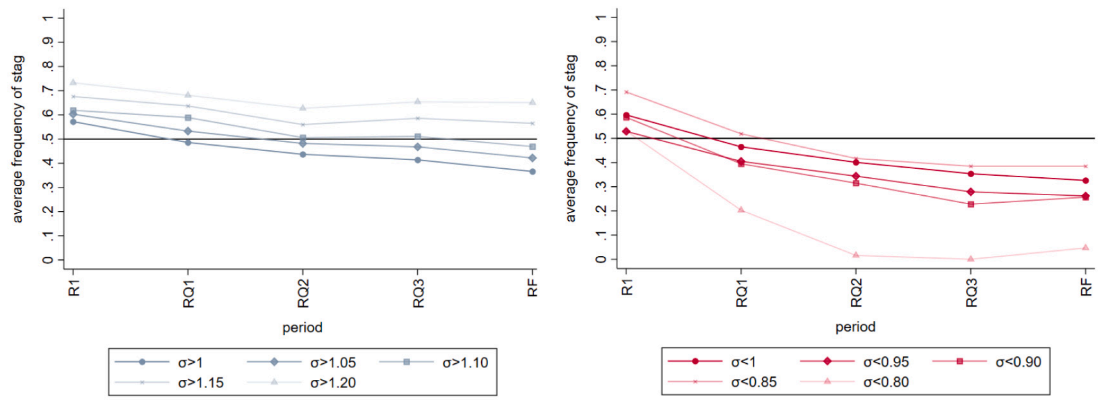

Fig. 8. Average frequency of stag when that option is more salient (left) and less salient (right) than hare.

functional specifications (see Table 8) where, instead of a ratio, we consider the difference between the salience of stag and hare (σ^-) or their product (σ^*). We do not find any of those alternative specifications to be superior to σ.

Table 4
Relative salience σ as a determinant of individual stag choices.

	(1) R1	(2) RQ1	(3) RQ2	(4) RQ3	(5) RF	(6) R1	(7) RF
σ	.446***	.750***	.761***	.906***	.824***	.688***	1.278***
	(.090)	(.145)	(.155)	(.164)	(.161)	(.188)	(.361)
σ^l						−.164	−.304
						(.114)	(.230)
obs	2528	2528	2528	2528	2528	2528	2528
pseudo R^2	.071	.110	.125	.168	.155	.072	.157
auc	.678	.718	.729	.763	.760	.679	.764
Hypothesis: $\beta_{\sigma,\mathrm{RF}} - \beta_{\sigma,\mathrm{R1}} = 0$							
chi2 (1)	9.79						
p-value	0.001						

Tables report average marginal effects in place of coefficients. Standard errors clustered at the session level in parenthesis. Every regression controls for article-specific effects, and the number of players in each session (which varies within an article). Feltovich et al. (2012) runs two different experiments (in the US and Japan) using different games. Hence, we treat each experiment as a different article. "auc" indicates the area under the curve. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table 5, we compare σ to other selection methods including the value components in Kendall (2022). We find that σ, except for choices in round 1 (possibly, due to some colinearity with other criteria, such as the basin of attraction) when we consider all methods simultaneously, accounts for the observed behavior, and thus, the probability of a subject choosing stag. Otherwise, we find that relative salience σ, when compared to another criterion (columns (1)–(8)), is always a significant determinant of stag choices at 1% level. In any case, in those regressions, we do find that *rr* (relative riskiness) emerges as a robust determinant of initial behavior (column (5)), and the *op* (optimization premium) is a weak determinant of behavior at later rounds (column (4)).

When we consider each criterion in isolation (Table 9 [Appendix 2]), the behavioral value *b*, as in Table 5, is still not a significant determinant of behavior. Given the statistical significance of σ, that result further supports our initial interpretation that the information missing from other selection criteria appears to be concentrated on the difference between each equilibrium's payoff, and not necessarily, as argued in Kendall (2022), the non-strategic information in that component. This finding also points against the conjecture in that same paper that information without an obvious strategic use (even as a coordination device) significantly affects the behavior in SHGs.

In Appendix 2, we re-estimate the regressions in Tables 4 and 5 at the session level (see Table 12 and Table 13). In those cases, the dependent variable is the frequency of stag choices in each session. The results are consistent with the evidence at the subject level.

One could eventually assume that, perhaps, it is not so much the trade-off between RISK and PAYOFF that explains the observed behavior in SHGs, but rather, the fact that σ accounts for subject-specific tastes and preferences (i.e., the random variable ε_i in Eq. (1)). For that reason, we re-estimated the regressions in Table 5 using Luce's relative salience (σ^l), which does not account for the idiosyncratic term ε_i (an option's unobservable attributes for a subject), instead of relative salience (σ). The results in Table 11 [Appendix 2] show that σ^l is also the only statistically significant determinant of stag choices at the end of a game compared to every other criterion. That result indicates that both the trade-off RISK and PAYOFF and unobservable subject-specific tastes account for the behavior in SHGs while confirming the importance of σ since it combines both aspects.

We move to Hypothesis 3 and look for differences in the relevance of σ as a determinant of behavior when either stag or hare are strongly or weakly salient. In that sense, we test four different thresholds: 5%, 10%, 15%, and 20%. We divide the sample into sessions where either stag or hare are strongly and weakly salient. When either of the options is strongly salient, as depicted in Table 6 [Panel A], σ is a significant determinant of stag choices at every threshold. There is also a particular trend in the correct classification of choices. For low thresholds (5% or 10%), the model does much better at classifying hare choices (i.e., predicting), reaching a success rate of 92%. However, as we increase the threshold τ, the model becomes much better at correctly classifying stag choices, reaching a success rate of 84% (with the lower fitting of hare choices most likely due to the low number of games where that option's salience is above the required threshold). Those results support our initial conjecture that convergence to stag requires larger levels of salience than hare.

In Table 7, we restrict the sample to games where the salience of hare is 5% or 10% larger than stag's, and games where the latter's salience is 15% or 20% larger than the former. We find that σ is a particularly good predictor of choices when stag's salience is 20% larger than hare's, and when the latter is 10% more salient than the former. In such cases, we correctly classify 84% of hare's choices and 72% of stag choices, resulting in a predictive gain of 24 p.p. against a rule where we expect subjects to pick stag in the first round and hare in the final one.

Additionally, consistent with Hypothesis 3, in Table 6 [Panel B], we observe that σ is not a strong determinant of subjects' choices when the salience of one option is just slightly larger (in that case, less than 5%) than the alternative. That picture changes when we increase the threshold to 10% but the marginal effect of a change in σ is not significantly different from the effect with a 5% threshold.

6. Conclusion

We propose a selection criterion motivated by the apparent tension between risk and efficiency in stag hunt games reported in numerous experiments. As a result, our criterion offers a relatively simple approach and explanation to the observed behavior in those games.

Table 5
Comparison between relative salience σ and other selection methods.

	(1) R1	(2) RF	(3) R1	(4) RF	(5) R1	(6) RF	(7) R1	(8) RF	(9) R1	(10) RF
σ	.529***	.972***	.433***	.769***	.413***	.788***	.528***	.978***	.300	.900**
	(.113)	(.195)	(.093)	(.163)	(.094)	(.169)	(.116)	(.204)	(.236)	(.398)
r	.142	.253							−.230	.719
	(.102)	(.196)							(.271)	(.546)
op			−.0001	−.0011*					−.00003	−.0059
			(.0001)	(.0006)					(.0001)	(.0045)
rr					.136**	.147			.225	−.149
					(.064)	(.128)			(.152)	(.252)
b							−.0016	−.0048	−.0002	−.0020
							(.0022)	(.0035)	(.0023)	(.0043)
s_1							−.0038	−.0071	−.0024	.0197
							(.0029)	(.0055)	(.0032)	(.0216)
s_2							.0014	.0004	.0012	.0015
							(.0022)	(.0041)	(.0023)	(.0043)
obs	2528	2528	2528	2528	2528	2528	2528	2528	2528	2528
pseudo R^2	.072	.157	.071	.160	.072	.157	.072	.159	.073	.164
auc	.678	.764	.678	.764	.680	.764	.680	.765	.679	.763

Tables report average marginal effects in place of coefficients. Standard errors clustered at the session level in parenthesis. Every regression controls for article-specific effects, and the number of players in each session (which varies within an article). Feltovich et al. (2012) runs two different experiments (in the US and Japan) using different games. Hence, we treat each experiment as a different article. σ: relative salience; r: hare's basin of attraction; op: optimization premium; rr: relative riskiness; b: behavioral value; s_1: stag's strategic value; s_2: hare's strategic value. "auc" indicates the area under the curve. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6
Relative salience σ as a determinant of individual stag choices in sessions where one option is strongly salient (panel A) or weakly salient (panel B) at some level τ.

Panel A: strong salience	$\tau = 5\%$		$\tau = 10\%$		$\tau = 15\%$		$\tau = 20\%$	
	R1	RF	R1	RF	R1	RF	R1	RF
σ	.502***	.970***	.317**	1.208***	.218*	1.080***	.218*	1.000***
	(.102)	(.203)	(.129)	(.287)	(.118)	(.278)	(.118)	(.263)
% correctly classified	.615	.730	.625	.730	.680	.708	.700	.728
% hare	.479	.913	.148	.916	.053	.814	0	.595
% stag	.709	.411	.929	.462	.975	.615	1	.840
gain	.039	.094	.014	.140	0	.243	0	.270
obs	1462	1462	988	988	588	568	356	356
pseudo R^2	.064	.182	.067	.211	.075	.215	.055	.270
auc	.659	.774	.657	.783	.676	.787	.653	.813
Panel B: weak salience	$\tau = 5\%$		$\tau = 10\%$		$\tau = 15\%$		$\tau = 20\%$	
	R1	RF	R1	RF	R1	RF	R1	RF
σ	.941	.811	.716***	1.168*	.668***	1.082***	.501***	.835***
	(.734)	(1.715)	(.238)	(.614)	(.185)	(.369)	(.150)	(.289)
obs	1066	1066	1540	1530	1940	1920	2172	2152
pseudo R^2	.083	.134	.080	.130	.068	.119	.071	.127
auc	.694	.742	.685	.740	.671	.728	.675	.737

Table reports average marginal effects in place of coefficients. Standard errors clustered at the session level in parenthesis. Every regression controls for article-specific effects, and the number of players in each session (which varies within an article). Feltovich et al. (2012) runs two different experiments (in the US and Japan) using different games. Hence, we treat each experiment as a different article. The "% correctly classified" indicates the percentage of correctly classified observations while "gain" indicates the excess in the number of correctly classified observations compared to a rule where every subject chooses stag in round 1 and hare in every subsequent round. "auc" indicates the area under the curve. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We find that subjects are more likely to pick stag the more salient is that option's defining aspect (its PAYOFF) compared to hare's main aspect (its RISK). As a game is repeated, the relative salience of each option seems to acquire additional importance, and the criterion appears to be relied upon more strongly as a coordination device. In particular, relative salience outperforms other selection methods (such as hare's basin of attraction, the optimization premium, relative riskiness, and different game components) as a determinant of the prevalence of stag. In any case, we note that the ability of this criterion to account for subjects' choices depends on the magnitude of each option's salience. In games where one option (e.g., stag) is not substantially more salient, other factors seem to acquire further importance in explaining the participants' beliefs and choices. In any case, our results are consistent with the conclusions in Dal Bó et al. (2021). Considering within-subject choices across sessions, they find that most subjects' behavior can be described as following a monotonic decision rule, such as "play stag if its basin of attraction is above a certain threshold". That rule, however, seems to be based, instead, on relative salience.

Dal Bó et al. (2021) also identifies differences in the prevalence of stag when subjects play one or multiple games for intermediate sizes of stag's basin of attraction. In our larger sample, we observe the same differences (Table 14 [Appendix 3]). However, we do not find

Table 7
Relative salience σ as a determinant of individual stag choices with different thresholds τ of strong salience for hare and stag.

	$\tau_{hare} = 5\%$ $\tau_{stag} = 20\%$		$\tau_{hare} = 10\%$ $\tau_{stag} = 20\%$		$\tau_{hare} = 5\%$ $\tau_{stag} = 15\%$		$\tau_{hare} = 10\%$ $\tau_{stag} = 15\%$	
	R1	RF	R1	RF	R1	RF	R1	RF
σ	.413***	.842***	.304**	1.128***	.463***	.919***	.317**	1.240***
	(.115)	(.212)	(.118)	(.246)	(.102)	(.189)	(.123)	(.273)
% correctly classified	.665	.768	.687	.784	.649	.745	.661	.750
% hare	.683	.835	.297	.840	.536	.859	.224	.875
% stag	.638	.668	.886	.717	.726	.573	.905	.594
gain	.064	.169	.025	.243	.054	.144	.020	.195
obs	818	818	568	568	958	958	708	708
pseudo R^2	.105	.303	.100	.348	.089	.260	.084	.283
auc	.708	.846	.709	.863	.694	.824	.687	.831

Table reports average marginal effects in place of coefficients. Standard errors clustered at the session level in parenthesis. Every regression controls for article-specific effects, and the number of players in each session (which varies within an article). Feltovich et al. (2012) runs two different experiments (in the US and Japan) using different games. Hence, we treat each experiment as a different article. The "% correctly classified" indicates the percentage of correctly classified observations while "gain" indicates the excess in the number of correctly classified observations compared to a rule where every subject chooses stag in round 1 and hare in every subsequent round. "auc" indicates the area under the curve. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

differences in the relevance of σ as a determinant of stag choices when subjects play one or multiple games each round (Table 15 [Appendix 3]). Overall, our results highlight the importance of a SHG structure when attempting to ensure efficient behavior (i.e., the choice of stag), and the necessity to ensure that stag's payoff is substantially more salient than hare's risk for that to happen.

CRediT authorship contribution statement

Rui Silva: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing.

Data availability

The authors do not have permission to share data.

Acknowledgments

I thank Françoise Forges, Paolo Garella, Mario Gilli, Marco LiCalzi, and Laurent Linnemer for their helpful comments. A special thanks go to Pedro Dal Bó, Russell Cooper, Douglas DeJong, Dimitri Dubois, Nick Feltovich, Ryan Kendall, Ismael Rodriguez-Lara, Thomas Ross, Larry Samuelson, David Schmidt, and Martin Sefton for kindly agreeing to share the data from their experiments.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.socec.2024.102290.

References

Battalio, R., Samuelson, L., & Van Huyck, J. (2001). Optimization incentives and coordination failure in laboratory stag hunt games. *Econometrica*, *69*(3), 749–764.
Candogan, O., Menache, I., Ozdaglar, A., & Parrilo, P. A. (2011). Flows and decompositions of games: Harmonic and potential games. *Mathematics of Operations Research*, *36*(3), 474–503.
Clark, K., Kay, S., & Sefton, M. (2001). When are Nash equilibria self-enforcing? An experimental analysis. *International Journal of Game Theory*, *29*(4), 495–515.
Cooper, R., DeJong, D. V., Forsythe, R., & Ross, T. W. (1992). Communication in coordination games. *Quarterly Journal of Economics*, *107*(2), 739–771.
Dal Bó, P., Fréchette, G. R., & Kim, J. (2021). The determinants of efficient behavior in coordination games. *Games and Economic Behavior*, *130*, 352–368.
Dubois, D., Willinger, M., & Van Nguyen, P. (2012). Optimization incentive and relative riskiness in experimental stag-hunt games. *International Journal of 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 and Economic Behavior*, *39*(1), 1–27.
Feltovich, N., Iwasaki, A., & Oda, S. H. (2012). Payoff levels, loss avoidance, and equilibrium selection in games with multiple equilibria: an experimental study. *Economic Inquiry*, *50*(4), 932–952.
Garcia-Galocha, A., Iñarra, E., & Iriberri, N. (2024). Do nonstrategic considerations matter for behavior in games? An experimental study informed by direct-sum decompositions of games. Available at SSRN: https://ssrn.com/abstract=4890257 or http://dx.doi.org/10.2139/ssrn.4890257.
Harsanyi, J., & Selten, R. (1988). *A general theory of equilibrium selection in games*. The MIT Press.
Kendall, R. (2022). Decomposing coordination failure in stag hunt games. *Experimental Economics*, *25*, 1109–1145.
Mengel, F. (2018). Risk and temptation: A meta-study on prisoner's dilemma games. *The Economic Journal*, *128*(616), 3182–3209.
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 and Economic Behavior*, *42*(2), 281–299.
Straub, P. G. (1995). Risk dominance and coordination failures in static games. *The Quarterly Review of Economics and Finance*, *35*(4), 339–363.