Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Politeness, conventionality and the evolution of social structures

Liping Tang

tanglp3@mail.sysu.edu.cn

Sun Yat-sen University

Research Article

Keywords:

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11751131/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Politeness, conventionality and the evolution of social structures

Abstract

Linguistic politeness is explained either as an individual strategy for mitigating social imposition or as a conventional practice acknowledging social relations. This paper brings these perspectives together through a graded notion of conventionality within evolutionary game theory. We develop a polite request game based on the stag hunt, distinguishing global reasoners, who interact uniformly, from local reasoners, whose interactions depend on social distance and power differences. We examine complete, ring, and star networks through stability analysis and simulations using fictitious play. The analysis identifies conditions favoring conventional or strategic politeness, while simulations frequently produce their coexistence. Conventional politeness generally predominates under fictitious play, whereas particular network configurations sustain strategic uses. These findings suggest that mixed patterns of politeness can emerge endogenously through the interaction of social structure, individual learning, and communicative coordination.

Keywords: Strategic and conventional politeness, Conventionality, Evolutionary stag hunt, Social structures

1 Introduction

Indirect speech acts are linguistic acts in which the speaker's meaning differs from the sentence meaning (Searle 1975). The following familiar example illustrates this distinction. A speaker making a request can choose between two utterances:

(1a) Pass me the salt!

(1b) Could you please pass me the salt?

Although (1b) conveys the same intention as (1a), it takes the more indirect form of a question. A speaker using (1b) generally expects the request to be fulfilled rather than answered with a simple yes or no. Why, then, does the speaker not use (1a) directly? Searle (1975) identifies politeness as the chief motivation for indirectness in such cases.

Two main perspectives explain the use of linguistic politeness (LP). Brown and Levinson (1987) explains politeness in terms of face-threatening acts: speakers use LP to mitigate the social imposition associated with requests. This account emphasizes

individual strategy. The conventional view, by contrast, emphasizes the social commitments underlying the evaluation of politeness (Matsumoto 1988; Ide 1989). Speakers use polite forms to acknowledge social relations. These views are compatible and can be situated on a continuum through the notion of conventionality (Terkourafi 2005, 2015; Simons and Zollman 2019; O'Connor 2021).

This paper examines how the notion of conventionality can formalize the combination of these two perspectives and how different degrees of conventionality emerge within an evolutionary game-theoretic framework.

First, we construct a polite request game based on the stag hunt, incorporating both "local" and "global" reasoners to represent strategic and conventional reasoning about LP, respectively. Second, we use evolutionary stability to measure conventionality, combining a population-level stability analysis with simulations of individual-level evolutionary dynamics. We also investigate how different social structures (complete, ring, and star networks) shape the emergence of different uses of LP. Our main findings are as follows.

1. The population-level stability analysis under best-response dynamics predicts convergence to either the conventional equilibrium or the strategic conventional equilibrium, with an evolutionary advantage for the former. A mixed equilibrium is rare.
2. In simulations using fictitious play across different parameter settings (population size, proportion of local reasoners, social structure, and memory size), the population often converges to a state resembling a mixed equilibrium.
3. When conventional and strategic conventional uses of LP coexist, conventional uses predominate in all simulations reported here, consistent with empirical findings in the literature (Shoshana et al. 1989; Chen et al. 2026a).

The remainder of the paper is organized as follows. Section 2 discusses the two views of LP and uses a game-theoretic account of conventionality to bring them together. Section 3 introduces the basic stag hunt game and develops a polite request game incorporating social structure. Section 4 analyzes the stability of this game under different network structures. Section 5 presents simulation results across parameter settings, focusing on convergence. Section 6 compares our model with related work. Section 7 concludes and outlines directions for future research.

2 Politeness and conventionality

2.1 Politeness with strategic and conventional explanations

Linguistic politeness (LP) is a central topic in pragmatics, both because it challenges Grice's cooperative principle (Grice 1975) and because it connects language use to social interaction (Goffman 1955), cross-cultural communication (Matsumoto 1988; Ide 1989), and the formation of social norms (Terkourafi 2005, 2015).

In their classic account, Brown and Levinson (1987) adopts the notion of "face" from Goffman (1955) to develop a general theory of politeness. Requests and demands are considered face-threatening acts. Politeness mitigates the imposition that these

acts place on interlocutors. Thus, the account in Brown and Levinson (1987) treats LP primarily as a matter of individual strategy.

In this account, politeness increases with the indirectness of an utterance. Its degree depends on social distance, differences in power, and the type of request. Although social relations play an important role in the theory of Brown and Levinson (1987), its explanation of LP remains grounded in the individual perspective.

Research on Japanese and Chinese has questioned whether politeness can be adequately explained as an individual's strategic avoidance of face threats, emphasizing instead the conventional use of LP (Terkourafi 2015; Matsumoto 1988; Yabuuchi 2006; Gu 1990). Polite forms, especially honorifics in these languages, acknowledge social relations: speakers use them to meet cultural expectations about behavior appropriate to their relative social position or rank.

Although these explanations emphasize individual strategy and social convention, respectively, they remain compatible.

> Consider the concepts of "collective, group culture" (East) and "individualist, egalitarian culture" (West). These are not absolute; they are positions on a scale. All polite communication implies that the speaker is taking account of both individual and group values. (Leech 2014, p. 83)

For example, *please* has been analyzed not only as face-threat mitigation but also as a conventional politeness marker whose function varies across English varieties (Murphy and Felice 2019). We argue that the two views can be synthesized in the same model via the notion of conventionality (Morgan 1978; Simons and Zollman 2019; O'Connor 2021; Tang 2024).

2.2 Measuring conventionality for indirect speech

Our discussion of conventionality draws on Lewis's account of convention (Lewis 1969). Following Lewis (1969), we call an element of meaning conventional if each speaker uses a given form with a given meaning only because others in the speech community do so. A convention emerges as a regularity in coordination rather than through negotiation and is characterized by arbitrariness. Although Lewis's standard definition treats convention as an all-or-nothing matter, he also briefly considers degrees of conventionality in terms of the expectation of participation in a regularity.

Subsequent work has developed this discussion. Millikan (2005) offers a historical and functional theory of convention that differs from Lewis's account but allows for gradations in conventionality. Building on Millikan's suggestion, O'Connor (2021) proposes an information-theoretic measure of conventionality. Although not explicitly framed in these terms, Terkourafi (2005) also allows for degrees of convention: an expression becomes conventionalized for a particular use in a context when, in a speaker's experience, it occurs frequently enough in that context to achieve a particular illocutionary goal.

Simons and Zollman (2019) argues that convention lies on a continuum from fully arbitrary to highly natural. Degrees of arbitrariness and naturalness thus describe opposite ends of the same scale. From a game-theoretic perspective, Simons and Zollman (2019) identifies three factors affecting a convention's arbitrariness or naturalness:

equilibrium payoffs, equilibrium stability, and availability. This framework also applies to indirect speech acts.

Drawing on the distinction in Clark (1991) between the means of making an indirect request and its linguistic form, Simons and Zollman (2019) shows that both dimensions vary in conventionality. As a request, *Can we open the windows?* is more conventional than *People sometimes open the window in here.*, illustrating variation in the conventionality of the means. The following examples illustrate variation in the conventionality of form.

> (a). Can you hand me that book on the top shelf?
> (b). Is it the case that you at present have the ability to hand me that book on the top shelf?
> (Searle 1975, p. 76)

Example (a) is more conventional than example (b). An unusual form typically suggests an unusual intention. Thus, both the means and the forms of indirect speech acts can be understood as conventions within a community. Their degrees of conventionality help explain the range of means and forms used to perform a given speech act.

Evolutionary game theory allows us to examine equilibrium payoffs and stability within a single framework. Coordination games are commonly used to characterize the emergence of conventions (Lewis 1969). In linguistic communication, successful coordination occurs when a speaker and a listener reach a shared understanding of meaning. An equilibrium thus represents successful coordination on a conventional meaning, and multiple equilibria represent alternative conventions.

	C_1	C_2
C_1	1	0
C_2	0	x

Table 1 Game 1: pure coordination $0 \leqslant x \leqslant 1$

	C_3	C_4
C_3	$3+y$	0
C_4	2	2

Table 2 Game 2: coordination with risk $0 \leqslant y \leqslant 1$

Tables 1 and 2 present a pure coordination game and a coordination game involving risk (a stag hunt), respectively. Each has two coordination equilibria: (C_1, C_1) and (C_2, C_2) in Game 1, and (C_3, C_3) and (C_4, C_4) in Game 2. In Game 1, when $x = 1$, the

equilibria have equal payoffs and stability, making the choice between them maximally arbitrary. When $x < 1$, (C_1, C_1) is both payoff-dominant and evolutionarily more stable, with a larger basin of attraction or greater resistance to mutant strategies (Young 1993). As x decreases, its advantage in both payoff and stability increases, making (C_1, C_1) a more natural convention. Thus, payoff and stability yield consistent assessments of conventionality in Game 1.

In Game 2, however, these criteria diverge. The equilibrium (C_3, C_3) offers a higher payoff, whereas (C_4, C_4) is more stable. As y increases, (C_3, C_3) becomes more stable. When y reaches 1, the two equilibria have equally large basins of attraction under certain dynamics of repeated play. When $y > 1$, (C_3, C_3) dominates in both payoff and stability.

We focus on Game 2 because it offers richer dynamics for investigating the evolutionary properties of conventions and their degrees of conventionality. Although (C_3, C_3) yields a higher payoff, (C_4, C_4) may be safer to play. A similar trade-off arises in communication: a familiar expression conveys information reliably, whereas an unusual expression may convey additional intentions and offer greater benefits, such as politeness appropriate to social relations, while risking miscoordination or misunderstanding.

In evolutionary game theory, the basin of attraction captures a feature of equilibrium stability that payoffs alone cannot: the range of initial population configurations that converge to an equilibrium. We use basin size to measure degrees of conventionality [1]. An equilibrium that attracts a larger share of initial configurations represents a more natural, or less arbitrary, convention.

When equilibria represent particular conventional uses of LP, a more natural convention corresponds to the strategic perspective, whereas a more arbitrary convention corresponds to the conventional perspective. We investigate how these uses of LP can be analyzed quantitatively using the communication pattern of Game 2 under different social structures.

2.3 Convention and social structures

Corpus evidence from Australian English, Argentinian Spanish, Canadian French, and Hebrew shows cross-linguistic variation in the conventionality of indirect speech forms and their illocutionary transparency (Shoshana et al. 1989). Conventionality therefore has cultural and social dimensions. Labov (1973) emphasizes that linguistic variation is socially organized: members of a speech community can differ systematically in their speech while sharing evaluative norms. Milroy and Milroy (1992) connects local interpersonal networks to broader class structures and emphasizes the role of loose or weak ties in facilitating linguistic innovation.

Research on Lewis's signaling games has extensively examined how structured communication shapes the emergence of meaning. Zollman (2005) shows that neighboring regions can develop different signaling conventions. Wagner (2009) demonstrates that communication systems depend strongly on network topology, with small-world graphs being especially conducive to the evolution of meaning. Mühlenbernd (2019)

[1] In the simulations in Section 5, this measure does not apply; instead, we use the long-run proportions of strategies to measure conventionality.

uses simulations of communicating agents in social networks to investigate environmental factors supporting semantic change or stability. Tang (2024) examines how social connections shape degrees of politeness.

Evolutionary game theory shows that structured interaction can substantially alter evolutionary dynamics. Under standard replicator dynamics with random matching, cooperation does not emerge in the prisoner's dilemma, whereas structured interaction can support it (Nowak 2006; Alexander 2007). In the stag hunt, structural effects depend on the dynamics: stag hunting performs well in two-dimensional populations under imitate-the-best dynamics, while hardly emerging under best response dynamics (Skyrms 2004).

To study politeness in social networks, we develop a network-based stag hunt version of the polite request game. We examine how the conventionality of the use of polite requests emerges through social dynamics in different network structures.

3 Stag hunt and conventionality

In a 2×2 stag hunt game, each of two players chooses between hunting a stag (S) and hunting a hare (H). If both hunt the stag, each receives the highest payoff, 3. If only one hunts the stag, that player receives nothing, while the hare hunter receives 2. Hunting the stag alone is therefore the worst outcome. Hunting a hare is safe: it yields a payoff of 2 regardless of the other player's choice. The payoff matrix is shown below.

	Stag (S)	Hare (H)
Stag (S)	3	0
Hare (H)	2	2

Table 3 Stag hunt

The game has two coordination Nash equilibria, (H, H) and (S, S). The former is risk-dominant and the latter is payoff-dominant (Harsanyi and Selten 1988). Although (S, S) yields greater social welfare, (H, H) has an evolutionary advantage because its basin of attraction is larger. When noise is introduced into the dynamics, (H, H) is also the stochastically stable equilibrium (Young 2020).

Let q denote the proportion of S players in the population. Using the payoffs in Table 3, a player is indifferent between the two actions when $3q = 2q + 2(1-q)$, giving the mixed-equilibrium threshold $q* = \frac{2}{3}$. Under best-response dynamics with random matching, players choose S in the next round when $q > \frac{2}{3}$ and H otherwise. The basins of attraction of (H, H) and (S, S) are therefore $\frac{2}{3}$ and $\frac{1}{3}$, respectively. In terms of initial strategy configurations, $\frac{2}{3}$ converge to (H, H), while $\frac{1}{3}$ converge to (S, S).

The stag hunt game has been used to study social coordination (Skyrms 2004; Young 1993) and linguistic conventionality (Simons and Zollman 2019).

3.1 Stag hunt for polite requests

Following the account in (Simons and Zollman 2019), conventionality lies on a continuum from fully arbitrary to highly natural. Naturalness can be assessed through equilibrium payoffs or stability. A higher-payoff equilibrium offers greater benefits, whereas a more stable equilibrium reduces the risks of coordination under uncertainty.

In Table 3, (S, S) is the more natural convention according to the payoff criterion, whereas (H, H) is more natural according to the stability criterion. The account in (Simons and Zollman 2019) leaves open how to assess naturalness when these criteria conflict, inviting further investigation of how natural conventions emerge within a community.

Consider the stag hunt game in Table 4. Let $r \in R^+$ represent the additional benefits associated with polite requests, with a higher r indicating a more difficult request [2]. Here, CP denotes "conventional politeness," and SP denotes "strategic politeness." Using CP, the established form for a particular request within a community, is generally safe. With SP, however, the listener may interpret only the literal meaning and miss the speaker's intention. In that case, the speaker receives the lowest payoff, 0, while the listener receives the safe payoff of 2.

	SP	CP
SP	$3 + r$	0
CP	2	2

Table 4 Stag hunt for polite requests

The mixed-equilibrium threshold is $q* = \frac{2}{3+r}$, so risk dominance can be assessed by comparing $\frac{2}{3+r}$ with $\frac{1}{2}$. If $\frac{2}{3+r} > \frac{1}{2}$, namely $0 < r < 1$, (SP, SP) is payoff-dominant and (CP, CP) is risk-dominant. If $\frac{2}{3+r} < \frac{1}{2}$, namely $r > 1$, (SP, SP) is both payoff-dominant and risk-dominant. Thus, when the benefits of a request are high, both criteria favor the natural convention (SP, SP). When benefits are relatively low, the criteria may favor different conventions.

Standard evolutionary dynamics favor (CP, CP), the risk-dominant equilibrium. This evolutionary advantage for conventional polite forms accords with empirical research on indirect speech (Shoshana et al. 1989). Languages differ in the conventionality of indirect speech forms and their illocutionary transparency. Nevertheless, conventional indirectness is more frequent than non-conventional indirectness or direct speech: it accounts for 82.4% in Australian English, 68.9% in French, 58.6% in Hebrew, and 58.4% in Spanish (Shoshana et al. 1989, p. 47). Strategic indirectness, termed non-conventional indirectness in Shoshana et al. (1989), nonetheless occurs at lower frequencies across languages. We therefore ask which evolutionary forces sustain it.

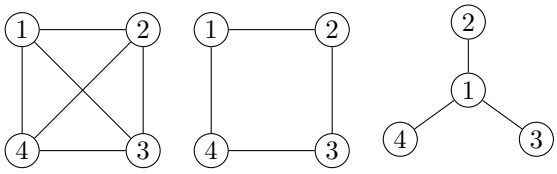
[2] We assume that strategic politeness involves greater cognitive costs but offers greater benefits when communication succeeds.

Dynamics, matching rules, initial configurations, and social structure may lead a community toward either equilibrium or a mixture of the two. We focus primarily on social structure while holding other parameters fixed.

3.2 Stag hunt for polite requests in networks

We investigate politeness under different social structures. The widespread use of honorifics in Japanese exemplifies conventional politeness, whereas strategic uses may be more common in English. These tendencies are matters of degree: both uses can coexist in different proportions (Leech 2014; Tang 2024). Social connections help shape these proportions (Brown and Levinson 1987).

Following the discussions in the literature (Zollman 2007; Tang 2024), we use three networks in our paper: completely connected, ring-shaped, and star-shaped networks in Figure 1.



(a) Complete (b) Ring (c) Star

Fig. 1 Three network structures with four players.

In a network $G = \{N, E\}$, N is the set of players, represented by nodes, and E is the set of undirected edges representing social connections. The social distance d_{ij} is the number of edges in the shortest path between players i and j. Player i's power is represented by the node's degree, deg_i, the number of edges incident to it.

According to Brown and Levinson (1987), social distance, power differences, and request type affect the degree of politeness in strategic polite requests. We incorporate these factors by modifying the game in Table 4: for speaker i and listener j in network $G = \{N, E\}$, where $i, j \in N$, we replace r with $r_{ij} = d_{ij} + |deg_i - deg_j| + r$.

	SP	CP
SP	$3 + r_{ij}$	0
CP	2	2

Table 5 Stag hunt for polite requests in G

The preceding analysis shows that a higher r_{ij} makes the strategic equilibrium more natural. Intuitively, successful communication offers greater benefits when a polite form mitigates a greater social imposition. Yet a higher r_{ij} and greater politeness also entail higher communication costs (Muehlenbernd et al. 2019). In the evolutionary

model developed below, we represent this disadvantage through lower matching probabilities: greater social imposition between players i and j reduces the probability that speaker i will choose listener j for a request.

4 Evolution and stability of conventions in social networks

The evolutionary properties of the stag hunt are well established. In Table 3, (H, H) is risk-dominant, corresponds to an evolutionarily stable strategy, and is stochastically stable under random matching (Young 1993; Kandori et al. 1993). Under best-response dynamics, the Pareto-inefficient equilibrium (H, H) can remain the long-run outcome even in structured populations, such as rings and lattices, where matching is not random (Ellison 1993; Skyrms 2004). The choice of dynamics matters, however: imitate-the-best dynamics can favor the payoff-dominant equilibrium in networks (Skyrms 2004).

We focus on best-response dynamics in networks and compare them with imitate-the-best dynamics in the simulations. Our principal departure from the literature is the following assumption.

In a population of size N, players are either global or local reasoners, with local reasoners constituting a proportion p. Their respective matching probabilities are π_{ij}^G and π_{ij}^L, defined as follows:

$$\pi_{ij}^G = \frac{1}{N-1}$$

$$\pi_{ij}^L = \frac{e^{-D_{ij}}}{\sum_{k \in N_i} e^{-D_{ik}}}$$

, in which $D_{ij} = d_{ij} + |deg_i - deg_j|$, N_i represents the set of i's neighbors excluding i.

Under π_{ij}^G, speaker i is equally likely to meet any other player. Global reasoners therefore treat all community members alike when choosing utterances, corresponding to conventional politeness, and use the payoffs in Table 4. Under π_{ij}^L, speakers select listeners according to their social relations, with greater social distance reducing the matching probability. Local reasoners thus correspond to strategic politeness and use the payoffs in Table 5. The adaptation rule incorporates both uses of polite forms.

These matching rules allow us to calculate each strategy's expected payoff for local and global reasoners. Because CP yields 2 regardless of the opponent's choice, $EU(CP) = 2$. We therefore need only calculate $EU(SP)$ under best-response dynamics.

For any network type $k \in \{com, ring, star\}$, player i's expected utility at t is calculated as follows. Let q_i denote the fraction of player i's opponents who play SP.

$$EU_i(SP)^k$$

$$= p \left(\sum_{j \in N_i} \pi_{ij}^L \big(q_i(3 + r_{ij}) + (1 - q_i) * 0 \big) \right)$$

$$
\begin{aligned}
&+ (1-p)\left(\sum_{j\in N-\{i\}} \pi_{ij}^{G}\big(q_i(3+r)+(1-q_i)*0\big)\right)\\
&= p\left(\sum_{j\in N_i} \pi_{ij}^{L}\big(q_i(3+r_{ij})+(1-q_i)*0\big)\right)\\
&\quad + \big(1-p\big)\big(q_i(3+r)+(1-q_i)*0\big)\\
&= p\left(\sum_{j\in N_i} \pi_{ij}^{L} q_i(3+r_{ij}))\right) + \big(1-p\big)q_i(3+r))\\
&= q_i\left(3+r+p\sum_{j\in N_i}\pi_{ij}^{L}D_{ij}^{k}\right)
\end{aligned}
$$

The second equality follows by substituting $\frac{1}{N-1}$ for π_{ij}^{G}. The final equality uses $r_{ij} = D_{ij}^{k} + r$.

If $EU_i(SP)^k \geqslant 2$, player i chooses SP; otherwise, the player switches to CP.

That is,

$$
\begin{aligned}
q_i\left(3+r+p\sum_{j\in N_i}\pi_{ij}^{L}D_{ij}^{k}\right) &\geqslant 2\\
q_i &\geqslant \frac{2}{\left(3+r+p\sum_{j\in N_i}\pi_{ij}^{L}D_{ij}^{k}\right)} \qquad (1)
\end{aligned}
$$

Let $\Delta_i^k = \sum_{j\in N_i}\pi_{ij}^{L}D_{ij}^{k}$ and then for all i and k, let

$$
\begin{aligned}
q^{\Delta} &= \frac{2}{\big(3+r+p\Delta_i^k\big)}\\
q_* &= \frac{2}{\big(3+r+p\max\Delta_i^k\big)}\\
q^* &= \frac{2}{\big(3+r+p\min\Delta_i^k\big)}
\end{aligned}
$$

Thus, we have the following results in terms of the population convergence result.

Proposition 1 *Let M^t be the number of players playing the payoff-dominant strategy SP at time t. Here, $[x]$ denotes the smallest integer greater than or equal to x.*

1. *If $M^t < [q_*(N-1)]$ at time t, then $q_i < q_*$ for every player i, and all players play CP in period $t+1$;*
2. *If $M^t > [q^*(N-1)]$ at time t, then $q_i > q^*$ for every player i, and all players play SP in period $t+1$;*
3. *If $M^t = [q^{\Delta}(N-1)]$ at time t, then players stay with their strategies at time $t+1$.*

Proposition 1 resembles the result in Ellison (1993) by providing sufficient conditions for convergence to either equilibrium. Convergence to the non-conventional polite form requires enough SP players for the threshold condition to hold simultaneously for every player. This is a demanding condition.

Under this condition, however, differences between network structures disappear. We therefore examine how these structural differences affect population stability.

Equation (1) shows that network differences enter through Δ^k. Holding other parameters fixed, higher Δ corresponds to higher r_{ij}. As the analysis of Table 5 indicates, this makes SP more likely to be selected; lower values favor CP.

For example, holding p, r, and n fixed, we can compare Δ across the networks in Figure 1.

- **Complete network** In the complete network, the distance between any two players is 1, and every player has the same degree; hence, $D_{ij} = 1$. Therefore, $\Delta_i^{com} = 1$ for any i.
- **Ring network** In the ring network, the power difference is also 0. $d_{12} = d_{14} = 1$, $d_{13} = d_{24} = 2$. Hence $D_{12} = D_{14} = 1$, $D_{13} = D_{24} = 2$. $\pi_{12}^L = \pi_{14}^L \approx 0.4223$, $\pi_{13}^L = \pi_{24}^L = 0.1554$, hence, $\Delta_i^{ring} = 1.1554$ for any i.
- **Star network**
  In the star network, we distinguish between speakers at the hub and those at the leaves.
  - When the speaker is the hub, then $d_{1j} = 1$; $|deg_1 - deg_j| = 2$. Hence, $D_{1j} = 3$. $\pi_{1j}^L = \frac{1}{3}$. $\Delta_1^{starH} = 3$.
  - When the speaker is at a leaf, we calculate Δ_2^{ring} for instance. $d_{21} = 1$, $d_{24} = d_{23} = 2$. $|deg_1 - deg_j| = 2$, $|deg_2 - deg_4| = |deg_2 - deg_3| = 0$. Therefore, $D_{21} = 3$. $D_{23} = D_{24} = 2$. $\Delta_2^{starL} = 2.1554$.

Thus, in the networks shown in Figure 1, star-network players have higher Δ values than players in the ring and complete networks. This result generalizes to N players as follows.

Proposition 2 *In networks with N players,*

$$\Delta_i^{starH} > \Delta_i^{starL} > \Delta_i^{ring} > \Delta_i^{com}, \quad i \in N > 3^{3}.$$

Proof See Appendix. □

Although a hub's network position offers an advantage under best-response dynamics, the evolutionary outcome depends crucially on the proportion of surrounding SP players. The hub has the highest social imposition value D_{ij} and consequently a low probability of being selected by others. For SP to persist or spread, the hub must play SP and the population must contain sufficiently many other SP players.

The stability analysis provides an overview of the game's evolution. At the individual level, however, dynamics, matching rules, initial configurations, and social

[3] When $N = 3$, the ring and the complete network coincide.

structure can make long-run population behavior difficult to predict. We explore these individual-level processes through simulations in the next section.

5 Evolutionary forces on conventionality in networks

Under myopic best-response dynamics, players respond to the current strategy distribution. Small changes around the threshold can trigger immediate cascades, making outcomes sensitive to the initial configuration and obscuring the effects of other evolutionary factors. To examine how individual behavior produces population-level outcomes, we use fictitious play, a widely used form of best-response learning. Each player responds optimally to the opponent's history of play. We focus on the long-run proportions of strategies rather than equilibrium stability.

The main purpose is to study the evolutionary outcome of this request game under different social structures. The polite request game simulated in this section has the following process:

1. Specify a population of N players in a complete, ring, or star network, together with $r \in (0, 1)$ and p.
2. Calculate π_{ij} and D_{ij} for each pair of players. Set $d = 0.1$ and let each neighbor contribute 0.1 to a node's degree [4].
3. Randomly assign $p * N$ players to be local reasoners and $(1 - p) * N$ to be global reasoners.
4. Randomly assign half the players to strategy S and half to strategy H to establish the initial configuration.
5. Repeat the game. At each time t, each player selects an opponent: local reasoners use matching probability π^L, whereas global reasoners use π^G.
6. Each player plays the stag hunt with the selected opponent and updates the choice of S or H through fictitious play, using the payoffs in Table 4 or Table 5. Local reasoners use $r_{ij} = D_{ij} + r$, whereas global reasoners use r.

We conduct a series of simulations, varying one parameter at a time. We investigate population size, the proportion p of local reasoners, and memory size under fictitious play, and compare fictitious play with imitate-the-best dynamics through networks.

5.1 Population size

We compare the effects of population size across complete, ring, and star networks. Initially, half the population plays SP and half plays CP. We fix $r = 0.5$ and $p = 0.5$, repeat the communication game for 50 rounds in each of 50 trials, and average the results. Figure 2 shows the average strategy proportions over time for populations of 5, 10, and 50.

First, we observe clear differences in the complete network. In the complete network, every $D_{ij} = 0.1$, and both local and global players meet every other player

[4] In Tables 4 and 5, an excessively large social-distance scale makes SP both risk-dominant and payoff-dominant, yielding a trivial result. We therefore control this scale so that CP can remain risk-dominant.

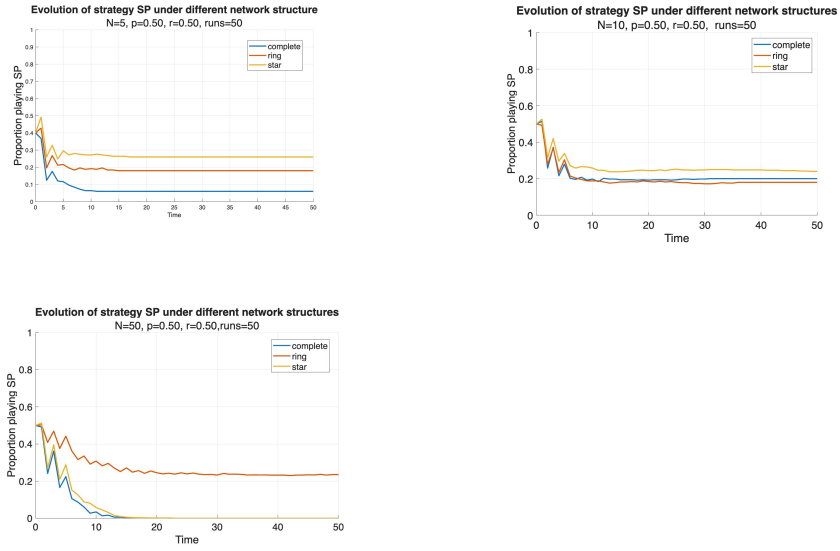


Fig. 2 Proportion of SP players in populations of 5, 10, and 50.

uniformly. However, they use different payoff thresholds:

$$q^{\Delta}_{\text{local}} = \frac{2}{3 + 0.5 + 0.1} \approx 0.5556,$$

whereas

$$q^{\Delta}_{\text{global}} = \frac{2}{3 + 0.5} \approx 0.5714.$$

The initial population contains 50% SP, below both thresholds. Therefore, following Proposition 1, the deterministic large-population tendency is toward H. This explains why the $N = 50$ simulation reaches 0% SP.

In the smaller population of 5 or 10 players, greater sampling noise can produce favorable early observations that push an individual's belief above the threshold. These individuals continue playing SP, generating further observations of SP for others and potentially sustaining a small group of SP players.

Second, the star network also varies with population size. As the population grows, π^L_{Hj} or π^L_{jH} decreases, and behavior becomes more uniform, approaching that of the complete network. The smaller star population therefore sustains a higher proportion of SP, whereas the larger population converges to 0%.

Finally, the ring network sustains a high proportion of SP players, mainly because q^{Δ}_{local} can fall below the initial proportion of 50%. As the network grows, D_{ij} increases and may exceed the threshold at which (CP, CP) remains risk-dominant. The equilibrium (SP, SP) can then become both payoff-dominant and risk-dominant. Consequently, SP remains common in the ring network at both population sizes.

5.2 Varying the weight of power differences in a two-star network

The hub plays a central role in the star network's evolution. In the current model, π^L_{Hj} is highly sensitive to population size. A large population yields a very small π^L for the hub player in the star-shaped network, thus leading the star network behavior to resemble that of the complete network. This is because, in a large star network, the power difference between the hub and leaf nodes is very large. We therefore introduce a weight on power differences: $D_{ij} = d_{ij} + \beta|deg_i - deg_j|, \beta \in (0, 1)$. To amplify the hub effect, we also consider a two-star network with two hubs. With a population of 10, we compare $\beta = 0.2$ and $\beta = 1$. Reducing the effect of power differences increases the hub's advantage for SP: the proportion of SP players rises from approximately 20% to 30% as β decreases from 1 to 0.2.

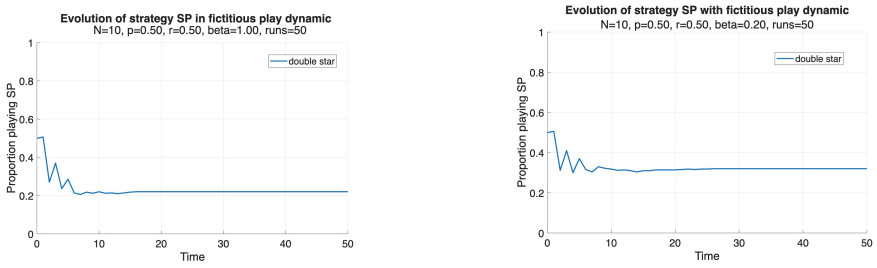


Fig. 3 Proportion of SP players in two-star networks: $\beta = 0.2$ vs. $\beta = 1$.

5.3 Varying p

Increasing the proportion of local reasoners increases the average D_{ij} relevant to play, particularly in the ring network, thereby increasing the proportion of SP. Figure 4 shows that, with a population of 50, the ring network is especially sensitive to changes in p.

5.4 Fictitious play with forgetting

In the preceding simulations, players were assumed to remember each individual's entire history of play. This is a demanding assumption. We therefore restrict memory to the previous 10, 5, or 3 rounds and compare the resulting population outcomes in Figure 5.

We rescaled D_{ij} to lie in $(0, 0.5)$. With $r = 0.5$, this ensures that $r_{ij} \in (0, 1)$ and removes the evolutionary advantage that SP gains when $r_{ij} > 1$. The resulting proportions of SP therefore reflect memory effects under this restriction. Figure 5 shows that limited memory affects the star network more strongly than the other two networks.

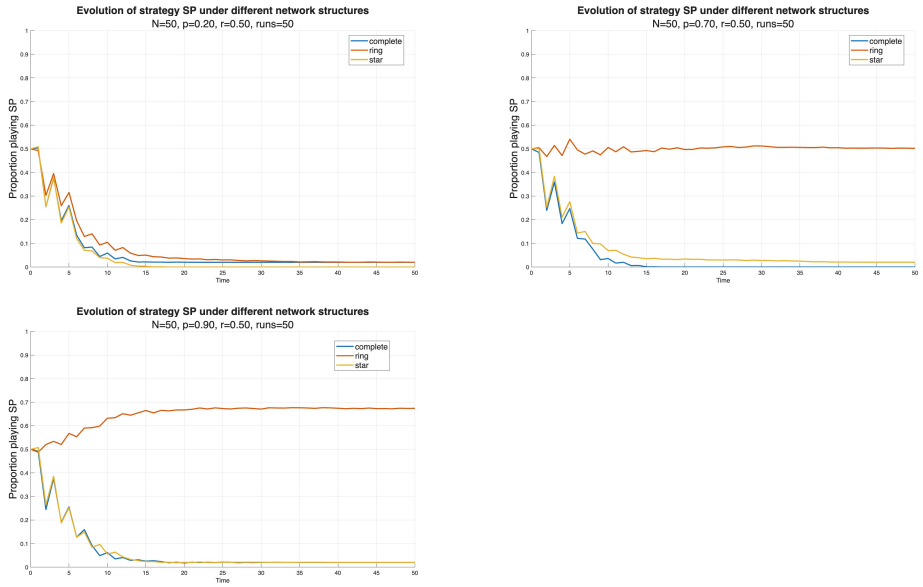


Fig. 4 Proportion of SP players with $p = 0.2, 0.7, 0.9$.

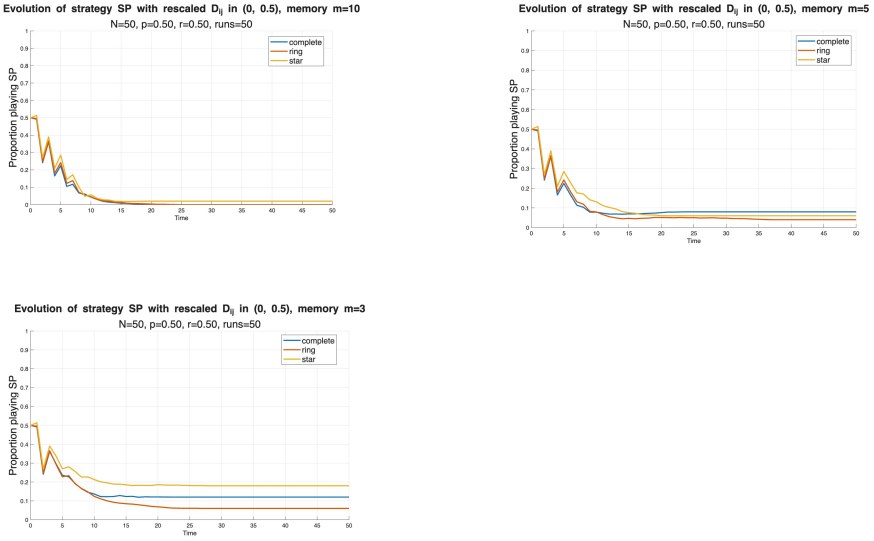


Fig. 5 Proportion of SP players with memory of 10, 5, and 3 previous rounds.

5.5 Fictitious play versus imitate-the-best dynamics

In the literature, imitate-the-best dynamics starting from an equal distribution of strategies lead the population to converge to stag hunting (Skyrms 2004). Figure 6 shows a similar result.

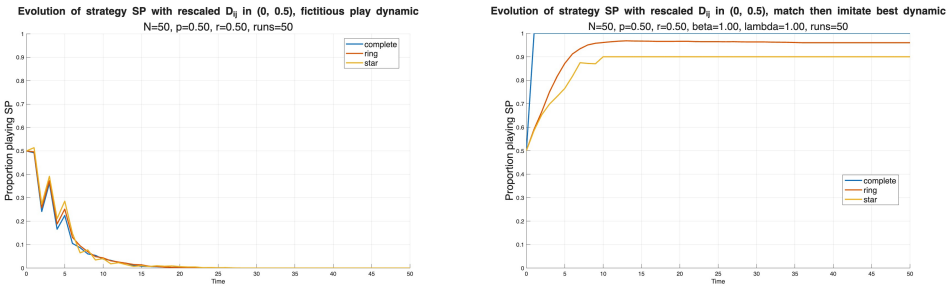


Fig. 6 Proportion of SP players under fictitious play and imitate-the-best dynamics

Figure 6 shows that, with rescaled D_{ij} and a population of 50, fictitious play leads to convergence to CP, the risk-dominant strategy, in every network. Under imitate-the-best dynamics, by contrast, the proportion playing the payoff-dominant strategy exceeds 90% in almost all networks. This comparison highlights the significance of the positive proportions of SP observed in the preceding fictitious-play simulations.

6 Discussion

The evolutionary stability analysis predicts the population (almost) converges to either a conventional polite strategy or a strategic polite strategy. The simulations, however, often produce long-run coexistence of the two strategies. Because the simulations more closely approximate actual communication, coexistence may also be more common in practice. This result is consistent with the literature and empirical findings (Leech 2014; Tang 2024; Chen et al. 2026b).

The effects of social structure on conventionality depend on parameter settings. In smaller populations with limited memory ($m = 3$), the star network strongly supports strategic politeness, consistent with Tang (2024). The ring network is sensitive to population size because larger populations increase social distance, favoring the strategic strategy and turning risk into a payoff advantage. In large star populations, however, the hub's advantage disappears: meeting probabilities decrease with power differences, making the hub less likely to be selected by its neighbors. Reducing the influence of power differences on π_{ij}^{L} can preserve the hub's advantage, as Figure 3 shows.

Although the two strategies coexist in most simulations, the strategic strategy is much less frequent than the conventional strategy under fictitious play. This accords with empirical observations that conventional indirect expressions occur more often across languages than strategic expressions (Shoshana et al. 1989). Artificial-language experiments likewise report more frequent conventional uses (Chen et al. 2026a).

Our paper differs from Tang (2024) in two main respects. First, that study assumes an exogenous mixture of conventional and strategic uses, introducing a weight $k \in (0, 1)$ into the payoff structure and examining how changes in k affect politeness across networks. Here, the mixture of polite uses emerges endogenously from interactions in different networks. Second, Tang (2024) assumes equal meeting probabilities and does not distinguish between local and global reasoners. Our model makes local reasoners' meeting probabilities depend on network structure.

7 Conclusion

We investigate the use of indirect speech acts, particularly polite requests, in a population through an evolutionary stag hunt game. The risk-dominant equilibrium captures the conventional use of polite forms as a safe mode of expression, whereas the payoff-dominant equilibrium captures their strategic use. Social structure shapes both the payoffs of the underlying game and the evolutionary dynamics. Our simulations show that the two strategies often coexist, with conventional use enjoying a long-run evolutionary advantage whose magnitude varies across network structures and parameter settings.

Statements and Declarations

Method and Data availability

We used AI assistants to polish author-written text, search for literature, and assist with simulation code. They did not generate any research questions, experimental designs, substantive interpretations, or final claims. The authors reviewed and edited the content as need and take full responsibility for the content of the article.

The main simulation code used in the model can be found through the link: https://osf.io/4n7hc/overview?view_only=496e458793c64f89ba91f2e4f1aff7bf

Competing interests

The author has no competing interests to declare that are relevant to the content of this article.

References

Alexander JM. The structural evolution of morality, vol. 10. Cambridge University Press Cambridge; 2007.

Brown P, Levinson SC. Politeness: Some universals in language usage, vol. 4. Cambridge university press; 1987.

Chen T, Yao Y, Tang L. Do Large Language Models Adapt Politeness Like Humans? In: Lee HY, Dev S, Li JJ, Elliott D, editors. Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing Budapest, Hungary: Association for Computational Linguistics; 2026. p. 15412–15429.

Chen T, Yao Y, Tang L. Do Large Language Models Adapt Politeness Like Humans? Proceedings of EMNLP2026. 2026;.

Clark HH. Responding to indirect speech acts. Pragmatics: A reader. 1991;p. 199–230.

Ellison G. Learning, local interaction, and coordination. Econometrica: Journal of the Econometric Society. 1993;p. 1047–1071.

Goffman E. On face-work: An analysis of ritual elements in social interaction. Psychiatry. 1955;18(3):213–231.

Grice HP. Logic and conversation. In: Speech acts. Brill; 1975. p. 41–58.

Gu Y. Politeness phenomena in modern Chinese. Journal of pragmatics. 1990;14(2):237–257.

Harsanyi JC, Selten R. A general theory of equilibrium selection in games. MIT Press Books. 1988;1.

Ide S. Formal forms and discernment: Two neglected aspects of universals of linguistic politeness. Multilingua. 1989;8(2-3):223–248.

Kandori M, Mailath GJ, Rob R. Learning, mutation, and long run equilibria in games. Econometrica: Journal of the Econometric Society. 1993;p. 29–56.

Labov W. Sociolinguistic patterns. No. 4, University of Pennsylvania press; 1973.

Leech GN. The pragmatics of politeness. Oxford University Press; 2014.

Lewis D. Convention. A Philosophical Study. Cambridge:Harvard University Press; 1969.

Matsumoto Y. Reexamination of the universality of face: Politeness phenomena in Japanese. Journal of pragmatics. 1988;12(4):403–426.

Millikan RG. Language: A biological model. Oxford University Press; 2005.

Milroy L, Milroy J. Social network and social class: Toward an integrated sociolinguistic model1. Language in society. 1992;21(1):1–26.

Morgan JL. Two types of convention in indirect speech acts. Syntax and semantics. 1978;9:261–280.

Muehlenbernd R, Zywiczynski P, Wacewicz S. The game theory of politeness in language: A formal model of polite requests. OSF. 2019;.

Mühlenbernd R. The change of signaling conventions in social networks. AI & SOCIETY. 2019;34(4):721–734.

Murphy ML, Felice RD. Routine politeness in American and British English requests: use and non-use of please. Journal of Politeness Research. 2019;15(1):77–100.

Nowak MA. Five rules for the evolution of cooperation. science. 2006;314(5805):1560–1563.

O'Connor C. Measuring conventionality. Australasian Journal of Philosophy. 2021;99(3):579–596.

Searle JR. Indirect speech acts. In: Speech acts. Brill; 1975. p. 59–82.

Shoshana BK, House J, Kasper G. Cross-cultural pragmatics: Requests and apologies. Grazer Linguistische Studien. 1989;.

Simons M, Zollman KJ. Natural conventions and indirect speech acts. Philosophers' Imprint. 2019;19(9):1–26.

Skyrms B. The stag hunt and the evolution of social structure. Cambridge University Press; 2004.

Tang L. Linguistic politeness in social networks. Synthese. 2024;203(6):204.

Terkourafi M. Beyond the micro-level in politeness research. Journal of Politeness Research: Language, Behavior, Culture. 2005;1(2):237.

Terkourafi M. Conventionalization: A new agenda for im/politeness research. Journal of Pragmatics. 2015;86:11–18.

Wagner E. Communication and structured correlation. Erkenntnis. 2009;71(3):377–393.

Yabuuchi A. Hierarchy politeness: What Brown and Levinson refused to see. Intercultural Pragmatics. 2006;.

Young HP. The evolution of conventions. Econometrica: Journal of the Econometric Society. 1993;p. 57–84.

Young HP. Individual strategy and social structure: An evolutionary theory of institutions. Princeton University Press; 2020.

Zollman KJ. Talking to neighbors: The evolution of regional meaning. Philosophy of Science. 2005;72(1):69–85.

Zollman KJ. The communication structure of epistemic communities. Philosophy of science. 2007;74(5):574–587.

Appendix

Proposition 2 *In networks with N players,*

$$\Delta_i^{starH} > \Delta_i^{starL} > \Delta_i^{ring} > \Delta_i^{com}, \quad i \in N > 3^{5}.$$

Proof See Appendix. □

Proof We calculate the $\Delta_i^k, k \in \{com, ring, star\}, i \in N$ for each network.

- Case 0: Obviously, $\Delta_i^{com} = 1$
- Case 1: a ring with an odd number of players $m = (N-1)/2$
  The weighted average local distance is

$$\Delta_i^{ring} = \frac{\sum_{k=1}^{m} ke^{-k}}{\sum_{k=1}^{m} e^{-k}}.$$

- Case 2: a ring with an even number of players $m = N/2$

$$\Delta_i^{ring} = \frac{2\sum_{k=1}^{m-1} ke^{-k} + me^{-m}}{2\sum_{k=1}^{m-1} e^{-k} + e^{-m}}, \qquad m = \frac{N}{2}.$$

- Case 3: the hub player in the star
  In the star network, if the player is the hub, all peripheral players have the same social distance from player 1, namely $N-1$. Hence,

$$\Delta_i^{StarH} = N - 1,$$

- Case 4: the player is a peripheral player in the star A peripheral player faces the hub at social distance $N-1$ and each of the other $N-2$ peripheral players at social distance two. Therefore,

$$\Delta_i^{StarP} = \frac{(N-1)e^{-(N-1)} + 2(N-2)e^{-2}}{e^{-(N-1)} + (N-2)e^{-2}}. \tag{2}$$

  For $N \geq 4$, Δ^{StarP} is a strict weighted average of 2 and smaller than $N-1$, so

$$2 < \Delta_i^{StarP} < N - 1 = \Delta_i^{StarH}.$$

  Moreover, the ring expressions are bounded above by the mean of the corresponding infinite geometrically weighted sequence,

$$\Delta_i^{ring} < \frac{\sum_{k=1}^{\infty} ke^{-k}}{\sum_{k=1}^{\infty} e^{-k}} = \frac{1}{1-e^{-1}} < 2.$$

[5] When $N = 3$, the ring and the complete network coincide.

It follows that, for $N > 3$,
$\Delta_i^{starH} > \Delta_i^{starL} > \Delta_i^{ring} > \Delta_i^{com}, \quad i \in N > 3$

□