Heuristics Facilitates the Evolution of Transitive Inference and Social Hierarchy in a Large Group

Kazuto Doi[1] · **Mayuko Nakamaru**[1]

Abstract

Transitive inference (TI) refers to social cognition that facilitates the discernment of unknown relationships between individuals using known relationships. It is extensively reported that TI evolves in animals living in a large group because TI could assess relative rank without deducing all dyadic relationships, which averts costly fights. The relationships in a large group become so complex that social cognition may not be developed adequately to handle such complexity. If members apply TI to all possible members in the group, TI requires extremely highly developed cognitive abilities especially in a large group. Instead of developing cognitive abilities significantly, animals may apply simplified TI we call reference TI in this study as heuristic approaches. The reference TI allows members to recognize and remember social interactions only among a set of reference members rather than all potential members. Our study assumes that information processes in the reference TI comprises (1) the number of reference members based on which individuals infer transitively, (2) the number of reference members shared by the same strategists, and (3) memory capacity. We examined how information processes evolve in a large group using evolutionary simulations in the hawk–dove game. Information processes with almost any numbers of reference members could evolve in a large group as long as the numbers of shared reference member are high because information from the others' experiences is shared. TI dominates immediate inference, which assesses relative rank on direct interactions, because TI could establish social hierarchy more rapidly applying information from others' experiences.

✉ Kazuto Doi
83nsoundbeach@gmail.com

Mayuko Nakamaru
nakamaru.m.aa@m.titech.ac.jp

1 Department of Technology and Innovation Management, Tokyo Institute of Technology, CIC-901S, 3-3-6 Shibaura, Minato-Ku, Tokyo 108-0023, Japan

1 Introduction

1.1 Background

How to increase chances of winning competitions for limited resources is critical for animals living in groups (Austad 1983; Enquist and Leimar 1983; Milinski and Parker 1991) because of significant impacts to the ability to survive and reproduce, or fitness. Previous studies on the evolution of animal conflicts examined which types of assessment of fighting ability, or resource-holding potential (RHP), of the opponent can evolve under different social conditions (e.g., Enquist and Leimar 1983; Hsu et al. 2006; Parker 1974; Reichert and Quinn 2017). Many early experimental studies assumed that body size, body mass, body color, voices and development of weaponry represent RHP. However, RHP is, sometimes, so invisible or intangible that sounds and colors are not useful at all, which is more often the case with the human society. When RHP is intangible, it is reasonable to assume that individuals can use outcomes from social interactions, such as wins or losses in animal contests, as indicators of RHP. Social interactions in animal contests play critical roles in forming the social hierarchy (Arnott and Elwood 2009; Hsu et al. 2006; Reichert and Quinn 2017).

There are many previous theoretical studies (e.g., Chase 1982; Dugatkin 1997, 2001; Dugatkin and Earley 2003; Nakamaru and Sasaki 2003) about types of assessment of RHP and the formation of the social hierarchy. Lindquist and Chase (2009) found that the winner-loser effect, defined as an increased probability of winning on past victories and an increased probability of losing on past defeats, does not show satisfactory agreement with the hen data they analyzed and suggested that individuals in a group are intensively aware of interactions among other members in their group.

Huang and Wu (2022) looked into the relationship between hierarchical structures and memory capacities and finds commonly observed hierarchical structures in nature as the total fitness-maximizing social structures given different levels of cognitive abilities.

Chase and Lindquist (2016) developed a theoretical approach that uses sequences of interactions with others within a group to explain the organization of the social hierarchy. They emphasized the importance of social cognition by taking eavesdropping, individual recognition and transitive inference as an example of social cognition (Hsu et al. 2006). Transitive inference is particularly important because it uses known relationships to deduce unknown ones. For example, A knows that A is stronger than B and B is stronger than C, but does not know if A is stronger than C. If A can have the ability of transitive inference, A can infer A > C, using A > B and B > C (Fig. 1a). Social cognition allows an individual to identify others, recognize and remember its relationship with others (Bshary and Brown 2014; Seyfarth and Cheney 2003).
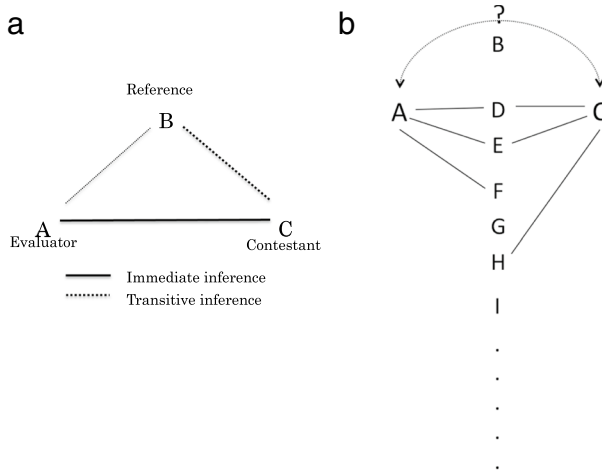
Fig. 1 **a** Transitive inference vs. immediate inference. Player A is an evaluator and will have a contest with player C (contestant). If A has direct contests with C in the past, A can evaluate C's strength based on the results of direct contests with C. This is called immediate inference. If A has no direct contests with C in the past but has direct contests with B (reference) who has direct contests with C, A can evaluate C's strength transitively based on the results of contests between A and B and the ones between B and C. This is called transitive inference. **b** Transitive inference process: the number of reference members and the number of shared reference members in TI_{X-Y} in the case of two TI_{3-2} players in a group. Players A and C follow the TI_{3-2} strategy. Given the number of reference members $= 3$, solid lines show that reference members for A are D, E and F and reference members for C, are D, E and H. Players applying the TI_{3-2} strategy are assumed to share two players with the other TI_{3-2} players since the number of shared reference members is 2. Shared reference members for A and C are D and E in the present example. Reference members are randomly chosen. Dotted line shows that A and C attempt to make an assessment of the relative rank each other using transitive inference-process when there are no direct contests between A and C. Player B adopts the immediate inference strategy

1.2 Transitive Inference and the Social Complexity Hypothesis

The social complexity hypothesis suggest that complex societies may promote the evolution of social cognition because members need to be able to handle the complexity from social interactions (Balda and Kamil 1989; Bond et al. 2003, 2010; de Waal and Tyack 2003; Jolly 1966). Reichert and Quinn (2017) and Hobson (2020) pointed out the importance of cognitive mechanisms that underlie contest behaviors, which little is known about.

Increasingly more studies support the social complexity hypothesis (Balda and Kamil 1989; Bond et al. 2003, 2010; Fernald 2014, 2017; Jolly 1966; MacLean et al. 2008; de Waal and Tyack 2003). As the group size increases, the number of possible interactions between pairs of individuals dramatically increases. Therefore, in a large group, it becomes increasingly difficult to understand the social hierarchy through the understanding of dyadic relationships between pairs of individuals in a group. For example, the number of members in a group in the study on the social hierarchies in *Astatotilapia burtoni* was 20 (Fernald 2014). In this case, every member has to remember the results of interaction with other all 19 members, which

takes a long time. This process takes a longer time as the group size increases. Transitive inference helps a member to assess RHP of the opponent the member has never interacted with before. Therefore, as the group size increases, transitive inference is more important. Transitive inference can assess opponents' RHP even without remembering all of the dyadic relationship with the opponents by using information from others. Therefore, transitive inference becomes increasingly important in the context of the social complexity hypothesis (Bond et al. 2003, 2010). Transitive inference is considered as a way of facilitating the understanding of social hierarchy without increasing the direct dyadic relationship under limited memory capacity (Mikolasch et al. 2013; Paz-Y-Miño et al. 2004).

Nakamaru and Sasaki (2003) and Doi and Nakamaru (2018) studied how transitive inference evolves, related to the social hierarchy formation theoretically by using the evolutionary game theory. These studies assume that individuals play the asymmetric hawk-dove game (Parker 1974; Maynard Smith 1974; Maynard Smith and Parker 1976). This framework has often been employed in the analysis of the evolution of fighting behaviors in animals.

Nakamaru and Sasaki (2003) demonstrated that the ability to accurately assess RHP is favored when the cost of losing is slightly larger than the benefit because the hawk vs. hawk combination occurs more often with lower costs. This ability is demonstrated in a strategy referred to as immediate inference (II) strategy in Doi and Nakamaru (2018) and Nakamaru and Sasaki (2003) where a player who estimates the strength of an opponent based on the history of direct fights or dyadic interactions (Fig. 1a). In contrast, the ability to form the social hierarchy promptly would be favored more when the cost is much higher than the benefit. This ability is associated with transitive inference (TI) strategy (Doi and Nakamaru 2018; Nakamaru and Sasaki 2003).

Both types of inferences have been reported extensively in the animal kingdom (Allen 2013; Grosenick et al. 2007; Paz-Y-Miño et al. 2004; Vasconcelos 2008; White and Gowan 2013). Both immediate and transitive inferences require social cognition, which refers to information learned about the characteristics of other individuals in the social interactions or based on observations (Sheehan and Bergman 2016). However, different types of inferences require considerably different types of social cognition. For example, immediate inference requires individuals to recognize only individuals that they have interacted with while transitive inference requires individuals to recognize an individual that they have never interacted with (Bshary and Brown 2014; Lilly et al. 2019; Seyfarth and Cheney 2015).

Social cognition has been investigated extensively in a wide range of animals, including both vertebrates and invertebrates (Emery et al. 2007; Gheusi et al. 1994). In particular, a recent report that transitive inference is observed even in insects such as wasps adds evidence that the miniature nervous system of insects does not limit sophisticated social behaviors (Tibbetts et al. 2019). This looks like a puzzle suggesting that the relationship between social information use in transitive inference and cognitive abilities may not be straightforward, given that social information use is common in taxa with advanced cognitive capacity like primates (Tibbetts et al. 2022). This study aims to show one of the possible answers to this puzzle.

Doi and Nakamaru ([2018](#)), which tried to answer the puzzle, theoretically demonstrated that transitive inference evolves with relatively low memory capacity when the cost of losing is relatively high because transitive inference can form the social hierarchy promptly. Furthermore, lower memory capacity is even more effective because lower memory capacity may restore the more consistent social hierarchy with true ranking based on actual RHP by disregarding existing social hierarchy that is not necessarily consistent with RHP.

However, Doi and Nakamaru ([2018](#)) still assumed highly developed social cognition that allowed individuals to recognize any other individuals that the individual had not interacted with and remembered all the outcomes of contests among any individuals. If all members apply transitive inference to all possible members in the group, transitive inference is supposed to require extremely highly developed cognitive abilities especially in a large group because every member needs to remember results of interactions by $N \times (N-1)/2$ pairs. Therefore, we relax this assumption in our present study. Our hypothesis in this study is that, instead of developing cognitive abilities substantially, animals may apply a kind of simplified transitive inference we call reference transitive inference as heuristic approaches. The reference transitive inference allows members to recognize and remember social interactions only among a set of reference member rather than all potential members. Animals may apply heuristic approaches such as reference transitive inference to handle such complex scenarios, instead of developing social cognition.

1.3 Heuristics, Collective Memory and Social Complexity

Individuals using reference transitive inference in a group infer the strength of unknown members transitively based on a set of reference members who can be a group of arbitrary members by applying information from experiences by other reference members. We also assume that the ability to share reference members with individuals following the same strategy (Fig. [1](#)b). Sharing reference members in a group requires not only cognitive capability but also social interactions such as some form of information exchanges among reference members, similarly to collective memories that refer to the pool of memories, knowledge and information shared in a social group. Coman et al. ([2016](#)) discusses the formation of collective memories in communication networks in laboratory-created communication. Weldon ([2000](#)) explored the formation of collective memories by using social network methodology. However, we did not consider the process to form shared memories and to choose reference members, for simplicity, in this study.

We redefine transitive inference as TI_{x-y}, where individuals can recognize and focus on an x number of reference members ($x \leq N-1$). Individuals following the same strategy share y number of members out of x number of reference members ($y \leq x \leq N-1$). Figure [1](#)b shows that how the number of reference members and the number of shared reference members interact in the transitive inference-process. Player A and C both employ a TI_{3-2} strategy. Considering the number of reference members $= 3$, we assume that the reference members for A are D, E and F, and the reference members for C, are D, E and H. Players applying the TI_{3-2} strategy are

assumed to share the information of two players D and E with other TI$_{3-2}$ players. D and E are shared reference members for all TI$_{3-2}$ players in the group. Shared reference members would help us to understand how sharing the same information in TI promotes the formation of the social hierarchy.

The heuristic approaches in reference transitive inference could substantially reduce the number of pairs required for understanding the entire social hierarchy, compared to immediate inference, when the group size is large. For example, immediate inference needs information about $N-1$ pairs and TI without heuristics requires information about $N \times (N-1)/2$ pairs, while TI$_{1-1}$ requires only $N-1$ relationships at minimum. Even limited number of reference members and shared reference members could facilitate the prompt establishment of the social hierarchy.

In this study, we consider the size of a group one of components influencing social complexity for simplicity. Social complexity is a common, but a little controversial, concept due to a lack of objectivity and a failure to link sociality to the application of cognition (Bergman and Beehner 2015). A review study about goldfish and parrots by Croney and Newberry (2007) and a comparative study of six primate species by MacLean et al. (2013) suggest that the group size signficantly influences the development of social cognition. However, the use of the group size as an index of social complexity is sometimes criticized because it does not take into account the diverse interactions among different animals within groups (Bergman and Beehner 2015).

2 Model

2.1 Hawk-Dove Game

We consider a group of N players. We pick two players, A and B, at random, from the group. A and B play the asymmetric hawk-dove game. Each player is supposed to select hawk (escalation) or dove (retreat). If both opt for dove, they share the reward V equally without fighting. Each receives $V/2$. If one player opts for hawk and the other player opts for dove, the hawk player wins and receives reward V. The dove loses and gains no reward. If both opt for hawk, they fight actually. The player who wins receives the reward V while the player who loses has to pay the cost, $-C$ ($V, C > 0$). The chance that A wins against B is determined by the difference of RHP of A and B based on the function $\theta(x_A, x_B)$ in the Eq. (1) below.

$$\theta(x_A, x_B) = \frac{1}{1 + e^{-(x_A - x_B)/a}} \tag{1}$$

where x_A and x_B correspond to RHP for players A and B respectively. Equation (1) suggests that when the A's RHP is higher than B's, A is more likely to win. When the value of a is lower, the probability that a player with a higher RHP would win is higher. The classical hawk–dove game assumes that $\theta(x_A, x_B)$ is 1/2 regardless of RHP.

It is an evolutionarily stable strategy (ESS) that players opt for hawk with a chance of V/C when $V < C$, or that players always opt for hawk when $V \geq C$.

2.2 Strategies and Assumptions

2.2.1 Three Types of Inference Processes

The strategies on which players select hawk or dove are genetically determined traits. We assume three types of strategies: mixer strategy (M), immediate inference strategy (II), and transitive inference strategy (TI_{x-y}) (Table 1).

We consider social cognition as a set of processes to (a) make an inference and (b) to gather and store the information for inference. The first part is referred to as inference processes while the second part is referred to as information processes (Table 1) in this study. As listed in Table 1, the strategies comprise of inference and information processes. Information processes comprise three parts: (1) The number of reference members who the focal individual can recognize and focus on, (2) The number of reference members shared by individuals (Fig. 1b), and (3) Memory capacity.

Table 1 Summary of strategies

	Inference processes			Information processes		
Strategies	TI-process	II-process	Mixer-process	x	y	MC
M	–	–	$\surd$	–	–	–
II	–	$\surd$(1)	$\surd$(2)	0	0	14
TI_{2-0}	$\surd$(2)	$\surd$(1)	$\surd$(3)	2	0	14
TI_{2-2}	$\surd$(2)	$\surd$(1)	$\surd$(3)	2	2	14
TI_{4-0}	$\surd$(2)	$\surd$(1)	$\surd$(3)	4	0	14
TI_{4-2}	$\surd$(2)	$\surd$(1)	$\surd$(3)	4	2	14
TI_{4-4}	$\surd$(2)	$\surd$(1)	$\surd$(3)	4	4	14
TI_{6-0}	$\surd$(2)	$\surd$(1)	$\surd$(3)	6	0	14
TI_{6-2}	$\surd$(2)	$\surd$(1)	$\surd$(3)	6	2	14
TI_{6-4}	$\surd$(2)	$\surd$(1)	$\surd$(3)	6	4	14
TI_{6-6}	$\surd$(2)	$\surd$(1)	$\surd$(3)	6	6	14
TI_{8-0}	$\surd$(2)	$\surd$(1)	$\surd$(3)	8	0	14
TI_{8-2}	$\surd$(2)	$\surd$(1)	$\surd$(3)	8	2	14
TI_{8-4}	$\surd$(2)	$\surd$(1)	$\surd$(3)	8	4	14
TI_{8-6}	$\surd$(2)	$\surd$(1)	$\surd$(3)	8	6	14
TI_{8-8}	$\surd$(2)	$\surd$(1)	$\surd$(3)	8	8	14

The mark $\surd$ shows which inference process each strategy adopts. The number in () next $\surd$ to indicates the order of priority in inference processes. For example, when (1) is available (1) is employed and when (1) is not available (2) is employed. 1 is the highest priority order and 3 is the lowest. MC in information processes stands for memory capacity defined as the number of contests players can remember

Each strategy consists of some of three types of inference processes, including mixer-process, immediate inference (II)-process, and transitive inference (TI)-process.

First, we will explain three types of inference processes. In mixer-process, a player makes a selection between hawk and dove following a mixed ESS where hawk is selected with a probability of V/C and dove with $1 − V/C$ when C (cost) $\geq V$ (reward). The player does not infer the strength of others. In addition, a player adopts the mixer-process when there is no assessment due to the lack of both ties and related contests.

We define R_X (B|A) as an assessment by player X of the relative rank of player B to A based on the past interactions between A and B. R_X (B|A) takes one of three values, $1, −1$ or 0. R_X (B|A) $= 1$ indicates that X assesses B stronger than A, if B has more wins than losses in the past contests between A and B. R_X (B|A) $= −1$ suggests that X assesses A stronger than B, if A has more wins than losses. R_X (B|A) $= 0$ means that X perceives A and B indifferent, if A ties with B or if there are no contests between A and B. We equally count as wins (losses) both wins (losses) in hawk vs. hawk and wins (losses) in hawk vs. dove. We consider only the signs, positive or negative, of differences of the numbers of wins and losses, not the magnitude of the differences.

In immediate inference-process, player A selects hawk when R_A (B|A) $= −1$ and dove when R_A (B|A) $= 1$. Similarly, player B opts for hawk when R_B (A|B) $= −1$ and dove when R_B (A|B) $= 1$.

With regard to the transitive inference-process, we assume that $TI_{x−y}$ players have the ability to observe and recall all contestants and results of contests only among x reference members in their information set where y reference members out of x are shared among the $TI_{x−y}$ players. The $x–y$ components in $TI_{x−y}$ represent the information processes characterized as a combination of two parts, (1) and (2) in information processes and correspond to heuristic mechanisms in transitive inference.

A set of shared reference members, referred to as y, is randomly determined from the group. Once y players are set, $(x–y)$ players are selected randomly from the group. We decide to select reference members randomly from the group, not in other ways. In fact, how to select reference members could depend on the relationships and availabilities among individuals under different social settings. Such realistic ways of selecting reference members would require more perplexing assumptions including various social contexts. Therefore, the random selection of reference members allows our study to focus on the complexity by the large group size.

Let us consider player A and B who need to assess the strengths each other. They have no direct contest, but both have direct contests with player C in the past. If R_A (B|C) $= 1$, or B > C, and R_A (C|A) $= 1$, or C > A, then R_A (B|A) $(= R_A$ (B|C) $+ R_A$ (C|A)) $= 2 > 0$, or B > A. If R_A (B|A) $(= R_A$ (B|C) $+ R_A$ (C|A)) > 0, then we set R_A (B|A) $= 1$. Here, transition inference suggests that if A < C and C < B, then A < B.

Similarly, if R_A (B|C) $= −1$, or C > B, and R_A (C|A) $= −1$, or A > C, then R_A (B|A) $(= R_A$ (B|C) $+ R_A$ (C|A)) $= −2 < 0$, or A > B. If R_A (B|A) $(= R_A$ (B|C) $+ R_A$ (C|A)) < 0,

then we set R_A (B|A) = -1. Then, transitive inference intimates that if A > C and C > B, then A > B.

If players cannot infer the strength of the opponent with transitive inference, the players follow a mixed ESS. For example, if R_A (B|C) = 1, or B > C and R_A (C|A) = -1, or A > C then A considers that B is as potent as A (R_A (B|A) = R_A (B|C) + R_A (C|A) = 0). If R_A (B|A) (= R_A (B|C) + R_A (C|A)) = 0, then we set R_A (B|A) = 0. In this case transitive inference suggests no difference between A and B.

We introduce a function $F(R)$ defined as follows to simplify the process: $F(R) = 1$ (if $R > 0$), $F(R) = 0$ (if $R = 0$), and $F(R) = -1$ (if $R < 0$). With the function, R_A (B|A) can be expressed as:

$$R_A(B|A) = F\big(R_A(B|C) + R_A(C|A)\big) \tag{2}$$

Generally, the number of opponents in common between A and B can be 2 or more. We refer to the individual common opponents as CO_i. We calculate $R_X(B|A)$ based on each CO_i. Then, transitive inference-process is defined as follows: CO_i are included in a set of players in the reference members and the maximum number of CO_i is x. The number of CO_i is n. Therefore, $R_X(B|A)$ can be expressed as:

$$R_X(B|A) = F\left(\frac{1}{n}\sum_i^n F(R_X\big(B|CO_i) + R_X\big(CO_i|A)\big)\right) \tag{3}$$

Using Fig. 1b, let us explain how player A and C, TI_{3-2} players, assess RHP each other. If A and C have direct contests with A's reference members, D, E and F, player A could assess the relative rank of A to C when there are no direct contests between A and C based on Eq. (3) as follows:

$$R_A(C|A) = F(1/3((F\big(R_A(C|D) + R_A(D|A)\big) + F\big(R_A(C|E) + R_A(E|A)\big)$$
$$+F\big(R_A(C|F) + R_A(F|A))))$$

If A does not have direct contests with F, R_A (F|A) is not available. The transitive inference-process is based on the following equation, instead of the equation above:

$$R_A(C|A) = F\big(1/2\big(F\big(R_A(C|D) + R_A(D|A)\big) + F\big(R_A(C|E) + R_A(E|A)\big)\big)\big)$$

Similarly, if C and A have direct contests with C's reference members, D, E and H, player C could assess the relative rank of C to A when there are no direct contests between the two based on Eq. (3) as follows:

$$R_C(A|C) = F(1/3((F\big(R_C(A|D) + R_C(D|C)\big) + F\big(R_C(A|E) + R_C(E|C)\big)$$
$$+F\big(R_C(A|H) + R_C(H|C)\big)))$$

Thus, the assessment by A of relative rank of A to C through shared reference members, D and E is common with the assessment by C of relative rank of C to A.

Therefore, the social hierarchies built by TI_{x-y} with more shared reference members will be more similar as the number of shared reference members increases.

Our assumption allows player D to be part of y players if D is also a TI_{3-2} strategist, because x and y are assumed to be selected from a group including the focal players. In this case, we define R_D (D|D) = 0. In general, R_X (X|X) is defined as zero when X represents a player employing the TI_{x-y} strategy.

In transitive inference-process, player A opts for hawk if R_A (B|A) < 0 and dove if R_A (B|A) > 0.

2.2.2 The Definition of the Strategies

The mixer strategy always employs mixer-process and does not require information about the contests. There are no reference members, no reference members shared by individuals, and no memory capacity in the information processes for the mixer strategy (Table 1).

Immediate inference strategy uses immediate inference-process basically and then mixer-process when the immediate inference-process does not produce information useful for an assessment based on information about contests the focal players directly involved. In the information process, there are no reference members and no reference members shared by individuals, and a limited memory capacity is needed for II-process (Table 1). TI_{x-y} strategy first relies on the immediate inference-process, shifts to the transitive inference-process when the immediate inference-process produces no useful information for an assessment and finally shifts to the mixer-process when the transitive inference-process results in no useful information (Table 1). Information processes for TI_{x-y} strategy are based on the contests by the x reference members where y reference members are shared. A memory capacity is needed for II-process and TI-process.

We focus on the situations where the group size, N, ranges from 10 to 50 members, large relative to the size of reference members and the cost of losing is high. This is because we consider the relative group size of cognitive abilities represented by the size of reference members, not absolute group size, is critically important in light of our research question. We defined the ranges of the number of reference members and the number of shared reference members both from 0 to 8 by 2 to facilitate the analysis of a broad range of parameters without a significant increase in computational complexity caused by an increase in the group size.

The present study employs 16 strategies in total; mixer, immediate inference, and 14 types of transitive inference strategies expressed as TI_{x-y}, including TI_{2-0}, TI_{2-2}, TI_{4-0}, TI_{4-2}, TI_{4-4}, TI_{6-0}, TI_{6-2}, TI_{6-4}, TI_{6-6}, TI_{8-0}, TI_{8-2}, TI_{8-4}, TI_{8-6} and TI_{8-8}. The strategies are designed to study how transitive inference evolves under the limited social cognition as defined above.

In our context, standard transitive inference, which appears in Doi and Nakamaru (2018) and Nakamaru and Sasaki (2003), is considered as TI_{N-N} when the group size is N. Standard transitive inference represents a unique case where the number of shared reference members, the number of reference members and the group size are all equal to N. In standard transitive inference, all players can recognize and recall all players and information about them in a group. Our study focuses on more

general circumstances with the number of shared reference members $\leq$ the number of reference members $<$ the group size, where players can recognize and recall only a limited number of other players in a group. TI_{x-y} represents more limited information processes than $TI_{N\text{-}N}$ because x and y are not greater than the group size, N.

When the group size is smaller and closer to the number of reference members, $(x–y)$ players are more likely to be overlapped among players with the same strategy TI_{x-y}. Before making a detailed explanation, our brief conclusion is that impacts should be very marginal when the group size is greater than 10 considering that the number of reference members is equivalent to eight.

Overlapping members in a set of reference members among the same strategists in the group emerges when the number of reference members is close or equal to the group size. When the number of reference members is equal to the group size, all members in the set of reference members are identical. Therefore, all members share all reference members ($x=y$ as a result). If a set of reference members is determined randomly from the group, assuming that the number of shared reference members is zero, we can count how many members in a set of reference members may overlap. As the number of reference members decreases to a level lower than the group size, the expected number of overlapped reference members among the same strategists, declines. For example, when the group size and the number of reference members are eight, any TI_{8-y} $(y<8)$ is identical to TI_{8-8}. When the group size is eight and the number of reference members is seven, the number of overlapped reference members declines substantially.

To clarify the impacts of the overlapping, we simulated how many reference members would overlap when the group size is ten assuming that a set of reference members is each determined randomly and the number of shared reference members is zero, or TI_{w-0}. We observe that the number of overlapped members among all members is 10 when $w=10$; four when $w=9$; one when $w=8$, and zero when $w=7$. These results suggest that such overlapping could influence TI_{8-y} $(y<8)$ marginally but would not affect any TI_{x-y} ($x \leq$ the number of reference members $=7$) when the group size is 10. Therefore, we do not consider that the overlapping could influence any TI_{x-y} when the group size is larger than 10. Overlapping would not matter overall because we focused on a large group.

2.3 Evolutionary Dynamics with Mutation

We consider a generation of T units of time. We assign a new RHP to each player at the beginning of each generation in a random manner and remains unchanged until it is reset. RHP is regarded as a nonhereditary trait expressed as a real number randomly chosen from a uniform distribution between 0 and 10, exclusive of 10. In one unit of time, two players who are randomly picked from the group play the hawk–dove game. The players opt for hawk or dove based on their strategies. After repeating the procedure T times in a single generation, the payoff for players is aggregated strategy by strategy. Subsequently, the number of players with the specific strategy at the start of the next generation is proportional to the aggregate payoff of players for the strategy in the prior generation. The aggregate payoff is

calculated to be positive by adding an absolute value of expected minimum payoffs to all players in order to avoid negative payoffs.

We assume that mutation occurs in the following two loci with a probability of μ independently: one is the number of reference members, referred to as x-locus and the other is the number of shared reference members, referred to as y-locus, which does not represent a set of specific members. A set of specific members is a nonhereditary trait.

Even though mixer and immediate inference strategies do not depend on the number of reference members or the number of shared reference members, we technically assign $x=0$ to the mixer strategy, $x=1$ to the immediate inference strategy, and $y=0$ to both mixer and immediate inference strategies. Then combinations of x and y are unique to each strategy so that mutation in the x and/or y loci means mutation in strategies.

We assume that mutation is allowed to occur randomly in the x-locus and then in the y-locus regardless of the current positions in the arrays. The new values in the x-locus and in the y-locus following mutation are allowed to adopt any values in the x-locus and the y-locus under $y \leq x$ conditions. So, $x \in \{0, 2, 4, 6, 8\}$. For each x, $y \in \{0\}$ in $x=0$, $y \in \{0, 2\}$ in $x=2$, $y \in \{0, 2, 4\}$ in $x=4$, $y \in \{0, 2, 4, 6\}$ in $x=6$, and $y \in \{0, 2, 4, 6, 8\}$ in $x=8$. For example, when the prevailing positions in x-locus and y-locus are 2 and 0, respectively, the new x-locus value following mutation could be 0, 4, 6 or 8, excluding 2, the current value, with the same probability, $\mu/4$. If the new value in the x-locus is 8, the new y-locus values could be 2, 4, 6 or 8 excluding 0, the prevailing value, with the same probability, $\mu/4$.

Finally, the next generation begins. This process continues over G generations. The group size is constant throughout a generation. Here we apply $\mu = 0.001$.

2.4 How to Measure the Social Hierarchy Formation

We introduce an analytical index modified based on consistency index (CI) developed in Doi and Nakamaru (2018) as an indicator of how consistency between $R_i(j|i)$ and $R_j(i|j)$ in any two players, i and j, evolves as players play games more, assuming that all players follow the same strategy in a group.

CI_1 measures how the number of reference members and the number of shared reference members influences the process of establishment of social hierarchy.

Details about CI are discussed in Doi and Nakamaru (2018). In short, $CI=0$ indicates that complete consensus is built where all combinations of tactics are hawk vs. dove or dove vs. hawk. Higher CI suggests more disagreements. The highest CI is 0.5, indicating complete disagreements.

In the present study we define CI_1 as $1 - CI/0.5$, where $CI_1 = 1$ indicates perfect consensus while $CI_1 = 0$ means no consensus. As players play games more and more, CI_1 ($0 \leq CI_1 \leq 1$) is expected to increase as a social hierarchy is established.

2.5 Key Parameters

There are four key parameters characterizing social conditions including (1) group size (N), (2) C/V ratio, which is a cost divided by a reward, (3) N_p ($= 2\ T/(N \times (N-1))$), referring to the expected number of contests per pair of players, and 4) Memory capacity (MC).

Here we use $N_p = 2$ because $N_p = 2$ gives two chances of participating in a contest to any pairs on average and Doi and Nakamaru (2018) suggest that TI works well under $N_p = 2$. $N_p = 2$ means that the encounter rates remain constant regardless of the group size because we increase T units of time as the group size increases. We use the constant $N_p = 2$ for all analyses in the present study for simplicity.

In the present study, we consider the group size (N) as one of components of social complexity as discussed in Sect. 1.2.

How reliable information from contests is in assessing RHP depends on the C/V ratio. For example, the probability ($= (V/C)^2$) that both players opt for hawk is low when C/V is high, so that results do not reflect actual RHP because the rank is often set without actual fights. The C/V ratio is a key parameter influencing what strategies can persist. We maintain the reward constant ($V = 4$) and vary the cost. We focus on the results when the cost is high ($C = 30$) because it is known that transitive inference persists in high-cost environments (Nakamaru and Sasaki 2003).

Memory capacity (MC) is defined as the number of contests players can remember. For example, immediate inference players maintain MC of records in memory about contestants and the results of their own direct contests. We assume that players forget older records beyond memory capacity and maintain only the latest MC of records. In the present study, we keep memory capacity constant ($MC = 14$) and then change group size in most of our analyses because we consider it reasonable to assume that memory capacity is limited (Miller 1956). The minimum memory capacity required for an individual to understand a relationship with others is $N - 1$. We consider $N - 1$ too low as a memory capacity; therefore, we set memory capacity as $2 \times (N - 1)$ given $N_p = 2$. $MC = 14$ assumes that the lowest size of a group is eight. When the group size is eight, $TI_{8\text{-}8}$ with $MC = 14$ represents adequate social cognition. This assumption means that individuals can remember 14 records of contests out of the expected numbers of encounters, 98 ($= 2 \times (50 - 1)$), when N is 50. All observations in memory are treated equally. In addition, we carried out sensitivity analysis of memory capacity by changing memory capacities under a constant group size.

However, inference process in the strategies gives priority to information about direct contests by first applying immediate inference-process, which is more direct experiences and then transitive inference-process in case of no direct contests.

3 Results

We explored the evolutionary dynamics of strategies in various group sizes, assuming that an initial strategy for all players is a mixer strategy. We ran the evolutionary simulations with mutation with all 16 strategies over 10,000 generations, iterated 50
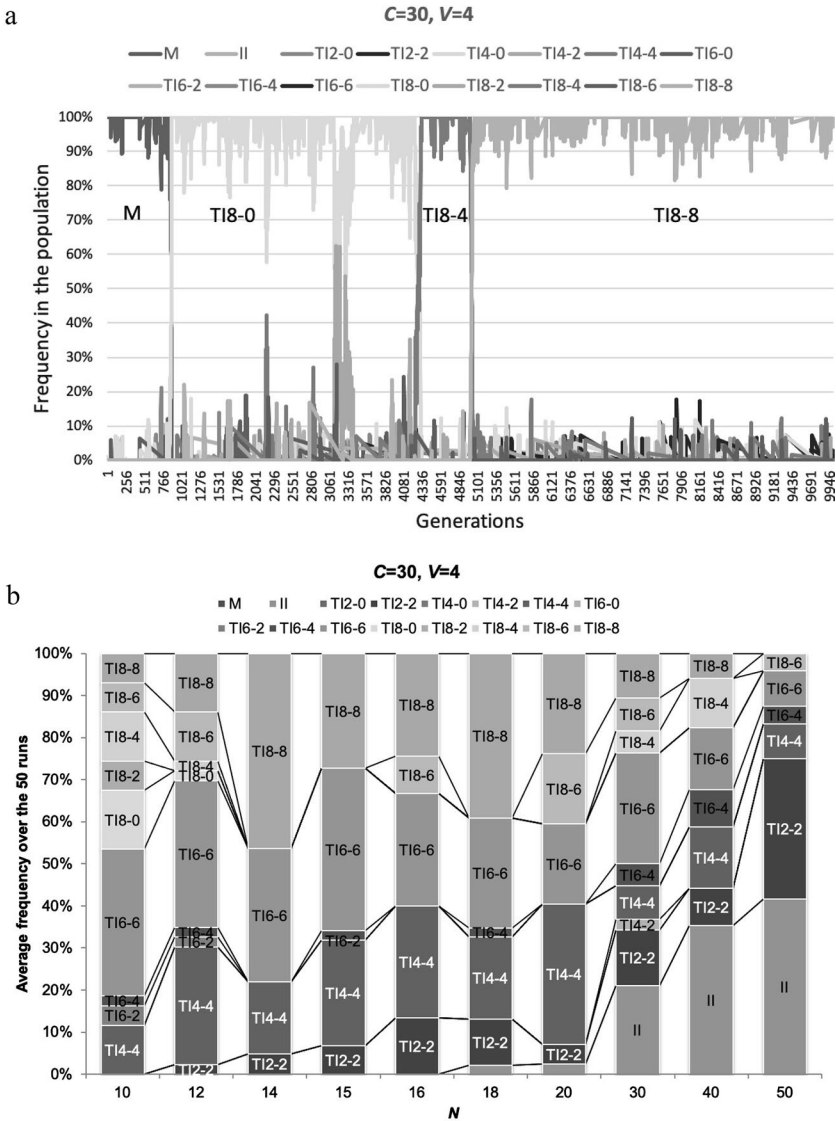
Fig. 2 Evolutionary simulation with random mutation. We examined evolutionary dynamics of all strategies with mutation that occurs in two loci with a probability of μ (=0.001) independently: one is for the number of reference members and the other is for the number of shared reference members. Initial strategy for all players is always a mixer strategy. Here we use the number of generations $(G) = 10\ 000$, $\mu = 0.001$, $MC = 14$, $Np = 2$, $V = 4$ and $C = 30$. In a, b, N varies under the constant MC. **a** The vertical axis represents the final frequencies of strategies over generations (G) in a single run and the horizontal axis represents G. Here we use $N = 20$. **b** The vertical axis represents the final frequencies of dominant strategies as averages over 50 iterations and the horizontal axis represents N. The vertical axis indicates how often each dominant strategy appears in the 50 iterations. **c** Here we use the number of generations $(G) = 10\ 000$, $\mu = 0.001$, $N = 14$, $Np = 2$, $V = 4$ and $C = 30$. MC (=52, 26, 18, 16, 14, 12, 10 and 8) varies. In c, MC varies under the constant N. The vertical axis represents the final frequencies of dominant strategies as averages over 40 iterations and the horizontal axis represents MC. The vertical axis indicates how often each dominant strategy appears in the 40 iterations
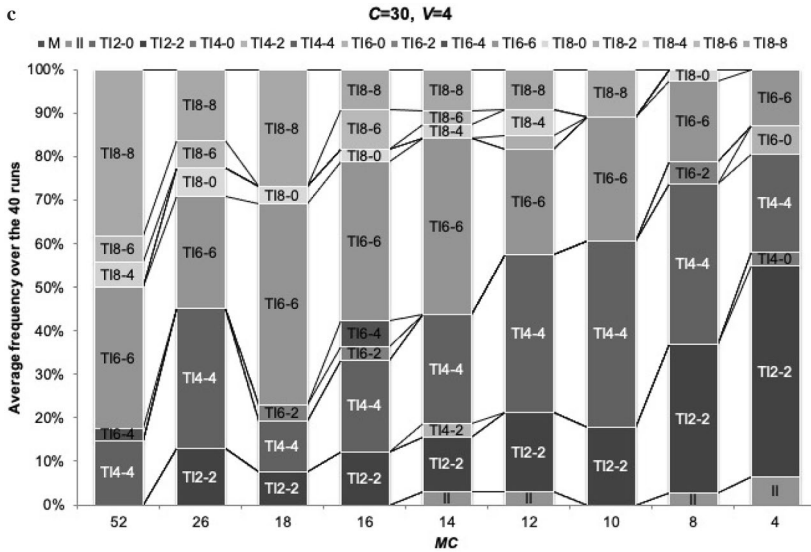
Fig. 2 (continued)

times, and calculated the average of population frequencies at each generation for each strategy. Each run out of 50 ends up with the dominant strategy without coexistence of strategies, as all strategies converges to a single strategy over the generations in a single run except a small number of mutants (Fig. 2a). A small number of mutants remains because mutation occurs randomly after one strategy dominates as designed.

Figure 2b shows the distribution of dominant strategies across different group sizes in the 50 iterations under the constant *MC*. Our analysis confirms that a group of reference transitive inference strategies is more dominant than the immediate inference strategy under limited cognitive abilities across any group sizes (Fig. 2b). More importantly, TI_{Z-Z} ($Z=2, 4, 6$ and 8) strategies turn out to be more successful than other TI strategies, TI_{Z-Y} ($Y < Z$). We also note that immediate inference strategy becomes more successful when $N \geq 30$. We will discuss the reason later in this section.

In addition, similarly to Fig. 2b, we looked into how different memory capacities (*MC*) impact the evolutionary dynamics of reference transitive strategies using the similar analysis applied in Fig. 2b where memory capacity (*MC*) is constant at 14 and the group size ($N = 10, 12, 14, 15, 16, 18, 20, 30, 40$ and 50) varies. In this *MC* sensitivity analysis, memory capacity ($MC = 52, 26, 18, 16, 14, 12, 10, 8$ and 4) varies and the group size (*N*) is constant at 14. It is, as expected, confirmed that a group of reference transitive inference strategies is more dominant than the immediate inference strategy under limited memory capacities and TI_{Z-Z} ($Z=2, 4, 6$ and 8) strategies dominate TI_{Z-Y} ($Y < Z$) (Fig. 2c). Hereafter, our analysis is based on the framework of different group sizes (*Ns*) under the constant memory capacity ($MC = 14$).
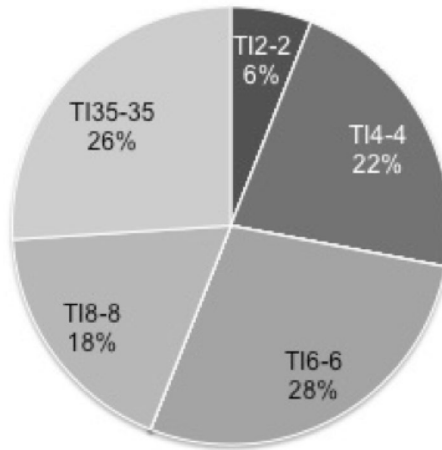
Fig. 3 Evolutionary simulation without random mutation. We analyzed evolutionary dynamics of M, II, TI_{2-2}, TI_{4-4}, TI_{6-6}, TI_{8-8} and TI_{35-35} with equal initial proportions under $N=35$. No mutation is assumed. We ran the process 50 times and calculated the average frequency of each strategy. Each run ended with 100% of the most dominant strategy and there was no coexistence of strategies. Final strategy frequencies represent how often the respective strategies become the most dominant strategy. We calculated the averages of the final frequencies only when the survival strategy converged into a single strategy. Here $G=500$, $MC=14$, $Np=2$, $V=4$ and $C=30$

Why, for example, is TI_{4-4} more successful than TI_{8-2} even though TI_{8-2} can recognize more group members? What if we introduce the standard transitive inference strategy, TI_{N-N} as TI_{N-N} can recognize all members. We looked into the evolutionary dynamics of strategies of M, II, TI_{2-2}, TI_{4-4}, TI_{6-6}, TI_{8-8} and TI_{35-35} under $N=35$ and $MC=14$. We assume that all strategies start with equal initial frequencies and no mutation occurs over 500 generations and repeated it 50 times. Similarly, each run ends up with dominant strategy without coexistence of strategies. Average final frequencies in Fig. 3 mean how often each dominant strategy appears in the 50 iterations.

Results in Fig. 3 show that final frequencies of TI_{4-4}, TI_{6-6}, TI_{8-8} and TI_{35-35} are similar, and TI_{2-2} ends up with the smaller final frequency. Immediate inference strategy does not survive in Fig. 3. It is noteworthy that TI_{35-35}, the standard transitive inference that can use information of all members in the group, does not make a meaningful difference from TI_{Z-Z} ($Z=4$, 6 and 8). It is true that TI_{Z-Z} ($Z=4$, 6, 8 and N is more successful than TI_{2-2} consistently both in Figs. 2b and 3. Results in Figs. 2b and 3 jointly suggest that the size of Z is not a key factor for survival as long as Z is greater than 2.

What is a key factor if the size of Z is not a key factor? Our hypothesis is that sharing more members (larger Y) is more critical than recognizing more members (larger Z) in forming the social hierarchy. By using the modified consistency index (CI_1) explained in the model section, we examined how consistently social hierarchies are built in TI_{Z-Z} and TI_{Z-Y} ($Y<Z$) under $Z=8$. Figure 4 demonstrates that TI_{Z-Z} strategies can form the social hierarchy better than TI_{Z-Y} ($Y<Z$) under any C. This result confirms our hypothesis that sharing reference members more with other
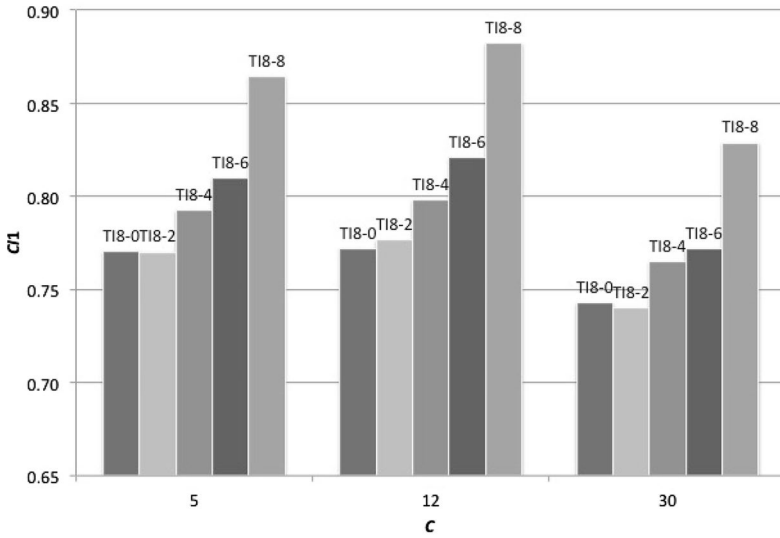
Fig. 4 *CI* level based on strategies with a constant number of reference members and different numbers of shared reference members. We ran 240 games in one generation with $N = 16$ ($Np = 2$) and $MC = 14$. The horizontal axis indicates C. The vertical axis represents average $CI1$ indices over 100 iterations. Here $V = 4$, $C = 5$, 12 and 30
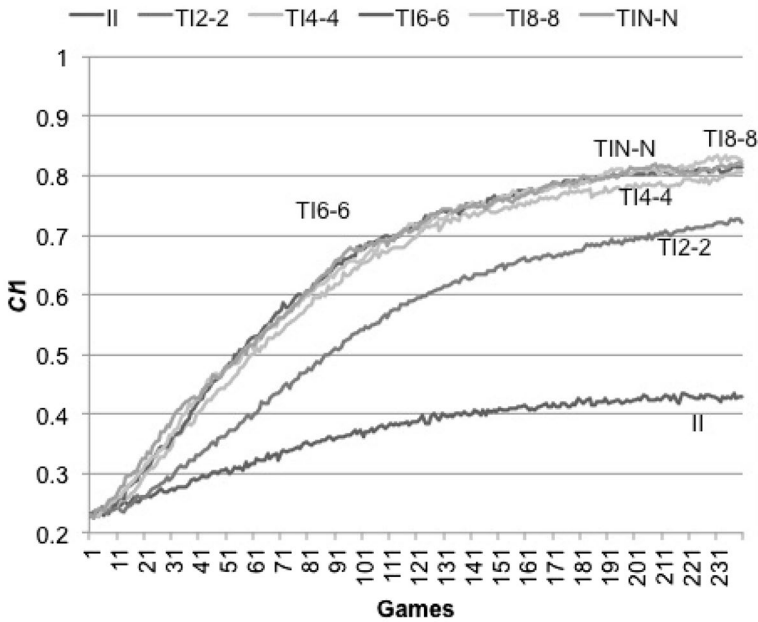

Fig. 5 *CI1* developments by strategies. We ran 240 games in one generation with $N = 16$ ($Np = 2$) and $MC = 14$. The horizontal axis indicates the number of games. The vertical axis represents averages of $CI1$ indices over 100 iterations. Line legend shows a strategy name. $C = 30$. Here $V = 4$

Fig. 6 $CI1$ developments in $\mathrm{TI}_{Z\text{-}Z}$ with unlimited memory capacity. We ran 210 games in one generation with $N = 15$ ($Np = 2$) under unlimited memory capacity. The horizontal axis indicates a strategy name. The vertical axis represents averages of $CI1$ indices over 50 iterations. $Z = 2, 4, 6$ and 8. $C = 30$. Here $V = 4$
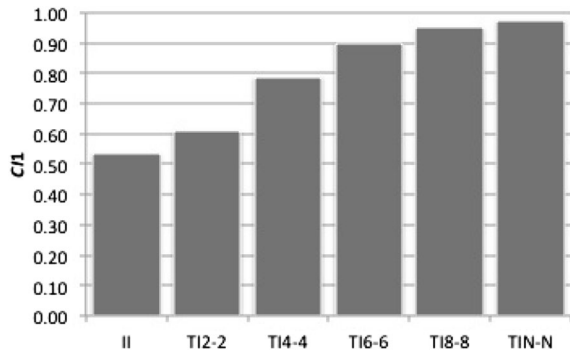
members promotes the prompt establishment of the social hierarchy by using information from others' experiences. The key factor to succeed is not the number of reference members (Z), but the number of shared reference members (Y) that is the ability to share the reference members (Fig. 4). In addition, Fig. 5 shows that $\mathrm{TI}_{Z\text{-}Z}$ ($Z = 2, 4, 6, 8$ and N) strategies can form the social hierarchy faster than immediate inference strategy. This promotes the evolution of transitive inference more in larger C where forming the linear social hierarchy is more important for survival.

Why does most $\mathrm{TI}_{Z\text{-}Z}$ strategies succeed similarly despite different Z? We examined how quickly the social hierarchies are built in II and $\mathrm{TI}_{Z\text{-}Z}$ ($Z = 2, 4, 6, 8$ and N) and found that CI_1 in all $\mathrm{TI}_{Z\text{-}Z}$ ($Z > 2$) also develops indifferently (Fig. 5). We consider that similar CI_1 behaviors are the reason for similar success in $\mathrm{TI}_{Z\text{-}Z}$ ($Z > 2$).

However, this finding seems a little counter-intuitive because higher Z should suggest higher cognitive abilities. We look into how CI_1 in all $\mathrm{TI}_{Z\text{-}Z}$ ($Z > 2$) develops under unlimited memory capacity and confirm that CI_1 with higher Z increases faster (Fig. 6). This means that limited memory capacity prevents $\mathrm{TI}_{Z\text{-}Z}$ with higher Z from being more successful. We find that all $\mathrm{TI}_{Z\text{-}Z}$ ($Z > 2$) including $\mathrm{TI}_{N\text{-}N}$, standard transitive inference, behaves similarly because of limited memory capacity. Z does not make a difference under restricted memory capacity relative to the group size.

Are $\mathrm{TI}_{Z\text{-}Z}$ strategies evolutionarily stable (ESS)? The evolutionary simulations beginning with all players applying $\mathrm{TI}_{8\text{-}8}$, $\mathrm{TI}_{4\text{-}4}$, or immediate inference end up with all players maintaining their respective strategies even at the end in a large group ($N = 40$) (Table 2). $\mathrm{TI}_{8\text{-}8}$, $\mathrm{TI}_{4\text{-}4}$ and immediate inference are all evolutionarily stable and could evolve if they are applied by the majority of a group. On the other hand, the evolutionary simulations beginning with all players applying $\mathrm{TI}_{8\text{-}0}$ or $\mathrm{TI}_{4\text{-}0}$ end with various combinations of final frequencies of different strategies (Table 2D and E). It is confirmed that $\mathrm{TI}_{8\text{-}0}$ and $\mathrm{TI}_{4\text{-}0}$ are not ESSs. In sum, $\mathrm{TI}_{Z\text{-}Z}$ ($Z <$ group size (N)) is an ESS while $\mathrm{TI}_{Z\text{-}0}$ is not an ESS because $\mathrm{TI}_{Z\text{-}Z}$ shares reference members with others while $\mathrm{TI}_{Z\text{-}0}$ does not. As discussed earlier, the ability to share reference members is critical because it facilitates the prompt establishment of the social hierarchy (Figs. 4 and 5).

Finally, why does immediate strategy start to appear again and $\mathrm{TI}_{Z\text{-}Z}$ with higher Z begins to dominate less when the group size becomes very large ($N \geq 30$) (Fig. 2b)? We consider that one of reasons is that the success of $\mathrm{TI}_{Z\text{-}Z}$ depends on initial

Table 2 Evolutionary dynamics of all strategies with the random mutations

	Frequencies	M	II	$TI_{2\text{-}0}$	$TI_{2\text{-}2}$	$TI_{4\text{-}0}$	$TI_{4\text{-}2}$	$TI_{4\text{-}4}$	$TI_{6\text{-}0}$	$TI_{6\text{-}2}$	$TI_{6\text{-}4}$	$TI_{6\text{-}6}$	$TI_{8\text{-}0}$	$TI_{8\text{-}2}$	$TI_{8\text{-}4}$	$TI_{8\text{-}6}$	$TI_{8\text{-}8}$
A	Initial																1.0
	Final																1.0
B	Initial							1.0									
	Final							1.0									
C	Initial		1.0														
	Final		1.0														
D	Initial					1.0											
	Final							0.7					0.1				0.1
E	Initial												1.0				
	Final							0.1					0.1			0.2	0.6

Evolutionary dynamics of all strategies with the random mutations that take place in two loci with a probability of μ $(=0.001)$ independently; one is for x and the other is for y in $TI_{x\text{-}y}$. Each case, A, B, C, D, and E has a different initial strategy frequency. Initial strategy frequencies are as follows; A with $TI_{8\text{-}8} = 100\%$, B with $TI_{4\text{-}4} = 100\%$, C with $II = 100\%$, D with $TI_{4\text{-}0} = 100\%$ and E with $TI_{8\text{-}0} = 100\%$. Numbers in each cell represent the strategy frequencies at the start (upper row) and the end (lower raw) for each case, as averages over 50 times. Each run ends up with 100% of the most dominant strategies and no coexistence of strategies. Final strategy frequencies represent how often the respective strategies become the most dominant strategy. We calculate an average of final frequencies only when the survival strategy converges into a single strategy. The numbers in cells are rounded and sum of the numbers may not be 1 because of the rounding. We examine cases with two different C/V ratios (1.25 and 4). Here we use $N = 40$, $G = 10\,000$, $\mu = 0.001$, $MC = 14$, $C = 30$ and $V = 4$

proportions of strategies. $TI_{Z\text{-}Z}$ with higher Z may require a higher initial proportion. We examined the evolutionary dynamics existing between immediate inference and $TI_{Z\text{-}Z}$ under different group sizes to observe how final frequencies of $TI_{Z\text{-}Z}$ develop over immediate inference with an increase in the group size (Fig. 7). No mutation was assumed. Figure 7 shows that $TI_{Z\text{-}Z}$ strategies with higher (lower) initial proportions tend to result in higher (lower) final frequencies. This result suggests that $TI_{Z\text{-}Z}$ has dependency on the initial proportions, meaning that $TI_{Z\text{-}Z}$ requires a larger number of players following the same strategy to recognize the similar hierarchy.

This dependency hypothesis is confirmed by results in Fig. 2c because immediate inference never emerges under the constant group size. In addition, $TI_{Z\text{-}Z}$ strategies survive over immediate inference in Fig. 3 where all strategies start with equal frequencies at the start, because $TI_{Z\text{-}Z}$ is more likely to survive when the initial frequency of $TI_{Z\text{-}Z}$ is greater (Fig. 7).

We consider that the other reason is that all strategies behave similar as N becomes very large ($N \geq 30$). For example, the $TI_{8\text{-}8}$ strategy with the largest cognitive abilities can observe, at the most, eight members in the group with more than 30 members. The $TI_{8\text{-}8}$ strategy would not make a difference from the $TI_{2\text{-}2}$ strategy that can observe only two members when the group size is very large. We looked into how CI_1 between $TI_{2\text{-}2}$ and $TI_{Z\text{-}Z}$ ($Z > 2$) develop under different N and find that CI_1 between $TI_{2\text{-}2}$ and $TI_{Z\text{-}Z}$ ($Z > 2$) becomes closer when N exceeds 30 (Fig. 8).

Fig. 7 Influence of initial population on TI$_{Z\text{-}Z}$. We analyzed evolutionary dynamics between II vs. TI with various initial proportions of TI$_{Z\text{-}Z}$ under $N = 30$. $Z = 2$, 4, 6 and 8. The vertical axis represents the final frequencies of TI $_{Z\text{-}Z}$ as averages over 50 iterations and the horizontal axis represents initial proportions of TI$_{Z\text{-}Z}$ as % share of an entire population. Here $MC = 14$, $Np = 2$, $V = 4$ and $C = 30$
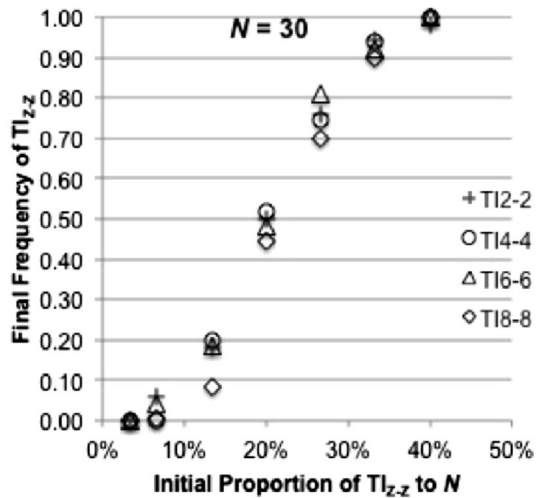
4 Discussion and Conclusions

We looked into three possible different attributes of cognitive capability: 1) How many group members an individual can recognize (reference members), 2) How many reference group members an individual can share (shared reference members) and 3) How many contests in a group an individual can remember (memory capacity). We assumed that all three are limited relative to the group size. Doi and Nakamaru (2018) theoretically studied the impact of limited memory to the evolutionary dynamics of transitive inference and the social hierarchy formation.

We have shown that a group of reference transitive inference strategies is more dominant than the immediate inference strategy under the constraints of limited cognitive capacities in a large group. This suggests that the ability in transitive inference to observe interactions among others, even though it is limited, distinguishes transitive inference from immediate inference because transitive inference can form the social hierarchy much faster than immediate inference as the group size increases.

The other important finding of our study is that, within the reference transitive strategies, TI$_{Z\text{-}Z}$ strategies dominate TI$_{Z\text{-}Y}$ ($Y < Z$) strategies and TI$_{Z\text{-}Z}$ strategies can persist even with the smallest Z across various group sizes (Fig. 2b and c). This is because sharing reference members more with other members (larger Y) promotes the prompt formation of the social hierarchy by using information from others' experiences. The ability to share reference members (larger Y) is more important than the ability to broaden a set of reference members (larger Z) especially when memory capacity is limited (Fig. 2b and c).

While prior research has discussed the importance and evidence of transitive inference in the formation of the social hierarchy especially in a large group (Hotta et al. 2015; Mikolasch et al. 2013; Paz-Y-Miño et al. 2004; Tibbetts et al. 2022), the question kept nearly unanswered on how such seemingly cognitively demanding task as transitive inference can succeed as group size increases. Our study proposes
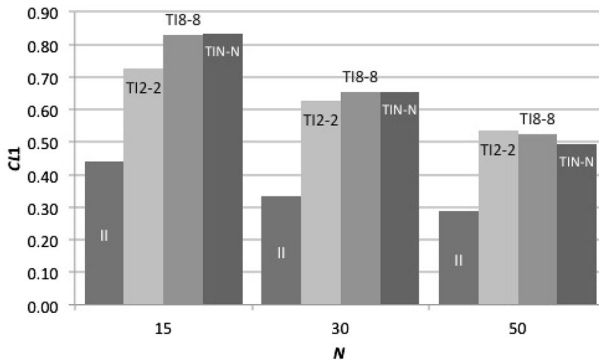
Fig. 8 $CI1$ developments in $TI_{Z\text{-}Z}$ under different group sizes. We ran 210 games in one generation with $N = 15$ ($Np = 2$) and with $MC = 14$. The horizontal axis indicates group size. The vertical axis represents averages of $CI1$ indices over 50 iterations for $N = 15$ and 30 and 30 iterations for $N = 50$. $Z = 2, 4, 6$ and 8. $C = 30$. Here $V = 4$

a theoretical solution to the question: Use of heuristics such as reference transitive inference where reference members share individual memories.

The present study reveals that reference transitive inference is not necessarily cognitively demanding in terms of the number of reference members. However, reference transitive inference can evolve in a large group. This observation is potentially inconsistent with the idea that more highly developed social cognition needs to evolve as the group size increases because the larger group size increases social complexity substantially. Our study suggests that animals may apply a type of short-cut, or heuristics, in order to deal with increasing social complexity with an increase in the group size, instead of developing very high levels of social cognition.

On the other hand, our study also shows that reference transitive inference requires the ability to share individual memories through shared reference members. The ability to share individual memories is part of cognitive capability but this should be also part of collective memory related to communication and a network in a group (Coman et al. 2016; Weldon 2000). In this present study, we assumed the ability to share individual memories, but did not discuss how this ability is developed, how this ability relates collective memory and how a network structure in a group impacts the formation of collective memory. These remains open questions left for future discussions about transitive inference.

Declarations

Conflict of interest Authors do not have any financial or non-financial interests that are directly or indi-

rectly related to the work submitted for publication.

References

Allen C (2013) Transitive inference in animals: reasoning or conditioned associations? Colin Allen, Texas A&M University. Behavioral Ecology, 1–16.

Arnott G, Elwood RW (2009) Assessment of fighting ability in animal contests. Anim Behav 77(5):991–1004. https://doi.org/10.1016/j.anbehav.2009.02.010

Austad SN (1983) A game theoretical interpretation of male combat in the bowl and doily spider (Frontinella pyramitela). Anim Behav 31(1):59–73. https://doi.org/10.1016/S0003-3472(83)80173-0

Balda RP, Kamil AC (1989) A comparative study of cache recovery by three corvid species. Anim Behav 38(3):486–495. https://doi.org/10.1016/S0003-3472(89)80041-7

Bergman TJ, Beehner JC (2015) Measuring social complexity. Anim Behav 103:203–209. https://doi.org/10.1016/j.anbehav.2015.02.018

Bond AB, Kamil AC, Balda RP (2003) Social complexity and transitive inference in corvids. Anim Behav 65(3):479–487. https://doi.org/10.1006/anbe.2003.2101

Bond AB, Wei CA, Kamil AC (2010) Cognitive representation in transitive inference: a comparison of four corvid species. Behav Proc 85(3):283–292. https://doi.org/10.1016/j.beproc.2010.08.003

Bshary R, Brown C (2014) Fish cognition. Curr Biol 24(19):R947–R950. https://doi.org/10.1016/j.cub.2014.08.043

Chase ID (1982) Dynamics of hierarchy formation: the sequential development of dominance relationships. Behaviour 80(3):218–239. https://doi.org/10.1163/156853982X00364

Chase ID, Lindquist WB (2016) The fragility of individual-based explanations of social hierarchies: a test using animal pecking orders. PLoS ONE 11(7):e0158900. https://doi.org/10.1371/journal.pone.0158900

Coman A, Momennejad I, Drach RD, Geana A (2016) Mnemonic convergence in social networks: the emergent properties of cognition at a collective level. Proc Natl Acad Sci USA 113(29):8171–8176. https://doi.org/10.1073/pnas.1525569113

Croney CC, Newberry RC (2007) Group size and cognitive processes. Appl Anim Behav Sci 103(3–4):215–228. https://doi.org/10.1016/j.applanim.2006.05.023

Doi K, Nakamaru M (2018) The coevolution of transitive inference and memory capacity in the hawk–dove game. J Theor Biol 456:91–107. https://doi.org/10.1016/j.jtbi.2018.08.002

Dugatkin LA (1997) Winner and loser effects and the structure of dominance hierarchies. Behav Ecol 8(6):583–587

Dugatkin LA (2001) Bystander effects and the structure of dominance hierarchies. Behav Ecol 12(3):348–352. https://doi.org/10.1093/beheco/12.3.348

Dugatkin LA, Earley RL (2003) Group fusion: the impact of winner, loser, and bystander effects on hierarchy formation in large groups. Behav Ecol 14(3):367–373. https://doi.org/10.1093/beheco/14.3.367

Emery NJ, Seed AM, von Bayern AM, Clayton NS (2007) Cognitive adaptations of social bonding in birds. Phil Trans R Soc B 362(1480):489–505. https://doi.org/10.1098/rstb.2006.1991

Enquist M, Leimar O (1983) Evolution of fighting behaviour: decision rules and assessment of relative strength. J Theor Biol 102(3):387–410. https://doi.org/10.1016/0022-5193(83)90376-4

Fernald RD (2014) Cognitive skills needed for social hierarchies. Cold Spring Harb Symp Quant Biol 79:229–236. https://doi.org/10.1101/sqb.2014.79.024752

Fernald RD (2017) Cognitive skills and the evolution of social systems. J Exp Biol 220(1):103–113. https://doi.org/10.1242/jeb.142430

Gheusi G, Bluthé R-M, Goodall G, Dantzer R (1994) Social and individual recognition in rodents: methodological aspects and neurobiological bases. Behav Proc 33(1–2):59–87. https://doi.org/10.1016/0376-6357(94)90060-4

Grosenick L, Clement TS, Fernald RD (2007) Fish can infer social rank by observation alone. Nature 445(7126):429–432. https://doi.org/10.1038/nature05511

Hobson EA (2020) Differences in social information are critical to understanding aggressive behavior in animal dominance hierarchies. Curr Opin Psychol 33:209–215. https://doi.org/10.1016/j.copsyc.2019.09.010

Hotta T, Takeyama T, Heg D, Awata S, Jordan LA, Kohda M (2015) The use of multiple sources of social information in contest behavior: testing the social cognitive abilities of a cichlid fish. Front Ecol Evol 3(August):1–9. https://doi.org/10.3389/fevo.2015.00085

Hsu Y, Earley RL, Wolf LL (2006) Modulation of aggressive behaviour by fighting experience: mechanisms and contest outcomes. Biol Rev Camb Philos Soc 81(1):33–74. https://doi.org/10.1017/S146479310500686X

Huang H, Wu J (2022) Limited cognitive abilities and dominance hierarchies. Acta Biotheor 70(3):1–19. https://doi.org/10.1007/s10441-022-09442-6

Jolly A (1966) Lemur social behavior and primate intelligence. Science 153(3735):501–506. https://doi.org/10.1126/science.153.3735.501

Lilly MV, Lucore EC, Tarvin KA (2019) Eavesdropping grey squirrels infer safety from bird chatter. PLoS ONE 14(9):4–8. https://doi.org/10.1371/journal.pone.0221279

Lindquist WB, Chase ID (2009) Data-based analysis of winner-loser models of hierarchy formation in animals. Bull Math Biol 71(3):556–584. https://doi.org/10.1007/s11538-008-9371-9

MacLean EL, Merritt DJ, Brannon EM (2008) Social complexity predicts transitive reasoning in prosimian primates. Anim Behav 76(2):479–486. https://doi.org/10.1016/j.anbehav.2008.01.025

MacLean EL, Sandel AA, Bray J, Oldenkamp RE, Reddy RB, Hare BA (2013) Group size predicts social but not nonsocial cognition in lemurs. PLoS ONE 8(6):e66359. https://doi.org/10.1371/journal.pone.0066359

Maynard Smith J (1974) The theory of games and the evolution of animal conflicts. J Theor Biol 47(1):209–221. https://doi.org/10.1016/0022-5193(74)90110-6

Maynard Smith J, Parker GA (1976) The logic of assymetric conflicts. Anim Behav 24(ii):159–175

Mikolasch S, Kotrschal K, Schloegl C (2013) Transitive inference in jackdaws (Corvus monedula). Behav Proc 92:113–117. https://doi.org/10.1016/j.beproc.2012.10.017

Milinski M, Parker GA (1991) Competition for resources. Behav Ecol. https://doi.org/10.1016/j.fcr.2009.06.017

Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63(2):81–97. https://doi.org/10.1037/h0043158

Nakamaru M, Sasaki A (2003) Can transitive inference evolve in animals playing the hawk-dove game? J Theor Biol 222(4):461–470. https://doi.org/10.1016/S0022-5193(03)00059-6

Parker GA (1974) Assessment strategy and the evolution of fighting behaviour. J Theor Biol 47(1):223–243. https://doi.org/10.1016/0022-5193(74)90111-8

Paz-y-Miño CG, Bond AB, Kamil AC, Balda RP (2004) Pinyon jays use transitive inference to predict social dominance. Nature 430(7001):778–781. https://doi.org/10.1038/nature02723

Reichert MS, Quinn JL (2017) Cognition in contests: mechanisms, ecology, and evolution. Trends Ecol Evol 32(10):773–785. https://doi.org/10.1016/j.tree.2017.07.003

Seyfarth RM, Cheney DL (2003) The structure of social knowledge in monkeys. In: Animal social complexity. Harvard University Press.

Seyfarth RM, Cheney DL (2015) Social cognition. Anim Behav 103:191–202. https://doi.org/10.1016/j.anbehav.2015.01.030

Sheehan MJ, Bergman TJ (2016) Is there an evolutionary trade-off between quality signaling and social recognition? Behav Ecol 27(1):2–13. https://doi.org/10.1093/beheco/arv109

Tibbetts EA, Agudelo J, Pandit S, Riojas J (2019) Transitive inference in polistes paper wasps. Biol Lett 15(5):20190015. https://doi.org/10.1098/rsbl.2019.0015

Tibbetts EA, Pardo-Sanchez J, Weise C (2022) The establishment and maintenance of dominance hierarchies. Phil Trans R Soc Lond B 377(1845):20200450. https://doi.org/10.1098/rstb.2020.0450

Vasconcelos M (2008) Transitive inference in non-human animals: an empirical and theoretical analysis. Behav Proc 78(3):313–334. https://doi.org/10.1016/j.beproc.2008.02.017

Waal FBM de, Tyack PL (2003) Animal social complexity; intelligence, culture, and individualized societies. Harvard University Press. http://www.hup.harvard.edu/catalog.php?isbn=9780674419131

Weldon MS (2000) Remembering as a social process. Psychology of learning and motivation. Academic Press, pp 67–120

White SL, Gowan C (2013) Brook trout use individual recognition and transitive inference to determine social rank. Behav Ecol 24(1):63–69. https://doi.org/10.1093/beheco/ars136