



A reinforcement learning model for grooming up the hierarchy in primates



Matthias Wubs^{a, *}, Redouan Bshary^b, Laurent Lehmann^a

^a Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

^b Department of Biology, University of Neuchâtel, Neuchâtel, Switzerland

ARTICLE INFO

Article history:

Received 12 July 2017

Initial acceptance 13 September 2017

Final acceptance 12 January 2018

MS. number: 17-00561R

Keywords:

food competition
grooming
learning
socioecological model
theoretical model
tolerance

Primates spend a considerable amount of time grooming each other. Grooming is regularly found to be traded reciprocally (for grooming) or for rank-related benefits in the presence of food competition. It has been suggested that if food sources are clustered and monopolizable, then lower-ranked individuals will groom higher-ranked ones in order to be tolerated on food patches. This leads to grooming being directed up the hierarchy. However, the conditions where this is expected to occur are based on verbal reasoning alone, and no quantitative analysis of the conditions favouring grooming up the hierarchy appear in the literature. Here, we develop a quantitative model investigating when food competition can result in grooming up the hierarchy. Individuals are assumed to take actions pertaining to whom to groom, where to feed and whom to tolerate on food patches. By allowing individuals to choose actions according to reinforcement learning, we delineate conditions where groups of individuals will express reciprocal grooming and grooming up the hierarchy depending on environmental conditions (e.g. quality, number of food patches). In particular, we show that conditions of intense food competition may lead to less grooming up the hierarchy. The predictions of our model could guide future comparative studies and meta-analyses investigating social relationships in primates.

© 2018 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

Primates spend a significant proportion of their daily time budget grooming other individuals: picking through fur to remove ectoparasites and clean the skin (social grooming, hereafter grooming; Dunbar, 1991). This observation may come as no surprise since being groomed is beneficial due to the removal of ectoparasites (Tanaka & Takefushi, 1993). Furthermore, grooming has an important social function, where social bonds between individuals are reinforced through the act of grooming. It has been argued that grooming behaviour has evolved because of these benefits (Dunbar, 1991). However, there can also be costs associated with giving grooming. These come in the form of opportunity costs (not being able to use the time or energy for other activities), such as reduced vigilance (Barrett & Henzi, 2006; Maestripieri, 1993). Thus, grooming behaviour can potentially be exploited, whereby within a pair of individuals only one makes the investment to groom and the other reaps the benefit without reciprocating. Understanding the causal proximate and ultimate factors influencing individual

grooming decisions is a long-standing goal in primatology (Schino, 2001; Seyfarth, 1980).

Grooming can be thought of as a commodity that can be traded, where individuals give grooming in order to receive something back. If grooming is solely traded for grooming, then pairs of individuals are expected to trade equal amounts of grooming. Numerous studies do indeed report a so-called time matching of grooming between individuals (Barrett, Henzi, Weingrill, Lycett, & Hill, 1999; Leinfelder, de Vries, Deleu, & Nelissen, 2001; Pazol & Cords, 2005; Rowell, Wilson, & Cords, 1991). Here, the individuals in a dyad are found to spend an approximately equal amount of time grooming each other, either within a single grooming bout or over a longer period of time. However, primate troops are nearly always characterized by a dominance hierarchy, and grooming interactions occur between individuals of different ranks. In a significant proportion of studies, grooming is observed to be directed up the hierarchy, that is, higher-ranked individuals receive more grooming than lower-ranked ones. Hence, grooming decisions are likely to depend on factors other than just the exchange of grooming. In other words, grooming of high-ranked by low-ranked individuals is likely to be traded for benefits other than grooming

* Correspondence: M. Wubs, Department of Ecology and Evolution, University of Lausanne, Lausanne 1015, Switzerland.

E-mail address: matthias.wubs@unil.ch (M. Wubs).

(Roubová, Konečná, Šmilauer, & Wallner, 2015; Seyfarth, 1980; Ventura, Majolo, Koyama, Hardie, & Schino, 2006; Wei et al., 2012).

What are the benefits that can explain grooming up the hierarchy? The socioecological model of primatology is concerned with the causes and consequences of food competition and its effect on social relationships (Kappeler & van Schaik, 2002). Within this framework it is argued that under conditions where individuals compete over food resources low-ranked individuals may trade grooming for nongrooming benefits provided by high-ranked ones (Kappeler & van Schaik, 2002; van Schaik, 1989; Sterck, Watts, & van Schaik, 1997). If food sources (or patches) are clumped and monopolizable, then individuals are expected to defend food sources or attempt to displace one another from them. Here, coalitions may be formed between individuals, either to defend a food source or to displace others from it. Then, low-ranking individuals may groom higher-ranking individuals to gain coalitional support in agonistic interactions (Carne, Wiper, & Semple, 2011; Schino, 2007; Seyfarth & Cheney, 1984; Seyfarth, 1980). Experimental evidence for grooming being traded for coalitional support can be found in baboons, *Papio hamadryas*, and vervet monkeys, *Chlorocebus pygerythrus* (Borgeaud & Bshary, 2015; Cheney, Moscovice, Heesen, Mundry, & Seyfarth, 2010). Owing to the trading of grooming by low-ranked individuals for such support from high-ranked individuals, on average grooming should be directed up the hierarchy.

Grooming can also be traded for increased tolerance (reduced aggression) for a low-ranked individual by a high-ranked one on a food source (Henzi & Barrett, 1999). This relies on broadly the same ecological conditions as the grooming under the coalitional support hypothesis. If food sources are monopolizable then, instead of attempting to displace a higher-ranked individual, the low-ranked individual may groom the higher-ranked one in order to be tolerated by that individual on the food source. Like the coalitional support hypothesis, there is empirical evidence showing that grooming for tolerance occurs (Carne et al., 2011; Tiddi, Aureli, Polizzi di Sorrentino, Janson, & Schino, 2011; Ventura et al., 2006; Xia, Garber, Sun, Zhu, & Sun, 2012). Furthermore, while most primate studies rely on correlational data, it has been shown experimentally that vervet monkeys trade grooming for short-term tolerance (Borgeaud & Bshary, 2015).

There thus exists ample empirical support that grooming may be directed up the hierarchy, when it is traded for either coalitional support or tolerance. However, the conditions under which this is expected to occur are mainly based on verbal models (van Schaik, 1989; Sterck et al., 1997; Wrangham, 1980), and there are few quantitative predictions as to the conditions under which grooming will be directed up the hierarchy and those grooming patterns that will be associated with it (which individuals receive extra grooming; Dunbar, 2002). One of the few models investigating grooming up the hierarchy is Seyfarth's (1977) model. In this quantitative model individuals make grooming decisions based on the 'attractiveness' of potential recipients. The attractiveness of an individual as grooming partner depends on the rank of that individual, since rank influences the likelihood of successfully supporting the focal individual in an agonistic interaction, and thus the highest-ranked individual is the most attractive grooming partner. The model managed to replicate two features that are typically found in various primate species: high-ranking individuals receive more grooming than low-ranking ones and individuals of adjacent rank groom each other more than expected by chance alone (Seyfarth, 1977). In an extension of this model, various types of competition for grooming access were investigated (Sambrook, Whiten, & Strum, 1995). However, both models explicitly assumed that high-ranked individuals are attractive grooming partners, and thus the grooming decisions are not linked directly to

the intensity of food competition itself. It would thus be relevant to have a quantitative model that links endogenous grooming decisions to the mode of food competition and that could determine conditions under which grooming up the hierarchy occurs as an equilibrium behaviour.

Here, we present a reinforcement learning (individual-based) model in which a group of individuals have grooming and feeding interactions, whose payoff consequences ('rewards') affect grooming, feeding and tolerance decisions. This allows us to determine conditions under which food competition can lead to grooming being directed up the hierarchy under the grooming for tolerance hypothesis. Although grooming occurs frequently between kin, a meta-analysis demonstrated a significant effect of grooming reciprocity among nonkin (Schino & Aureli, 2008). Since our main interest here is the effect of food competition, we assume that individuals in the model are unrelated. However, we further assume that individuals spend their entire life together, which corresponds to the philopatric sex, and this typically means the females in primates (Pusey & Packer, 1987). The main questions that we address with this model are: (1) can food competition result in grooming up the hierarchy as an equilibrium behaviour when individuals decide whom to groom, where to feed and whom to evict, dynamically according to payoffs, (2) what kind of grooming patterns are expected if grooming is directed up the hierarchy (e.g. do all individuals groom the alpha individual, or do they mostly groom one rank up the hierarchy), and (3) to what extent does the strength of competition (due to the number of patches or the aggressiveness of individuals) modulate grooming up the hierarchy?

THE MODEL

The Biological Setting

We consider a group of N individuals ranked in a stable, linear dominance hierarchy where individual $i \in \{1, 2, \dots, N\}$ has a unique rank $r_i = i \in \{1, 2, \dots, N\}$. Individual $i = 1$ is the lowest-ranked and $i = N$ is the highest-ranked or alpha individual. Individuals within this group interact with each other for T discrete time steps, which can be thought of as the number of interactions on a daily basis, monthly basis, yearly basis, etc. Each interaction time step $t = 1, 2, \dots, T$ is characterized by four sequential behavioural stages: grooming, food patch selection, aggressive interactions and feeding, which occur in this order and that we now detail.

Stage 1: Grooming

Each individual either selects an individual that it grooms or it grooms no one. The set of actions available to individual i in this

Table 1
List of model parameters

Parameter	Equations	Meaning
c	6	Cost of grooming
b	6, 10	Benefit of being groomed
γ	6	Change in motivation when not grooming
λ	4, 8	Exploration rate
δ	5, 9, 10	Discount factor
β	11	Regulate probability of engaging in a contest
ϵ	10	Impact of eviction
d	1, 7	Dominance gradient
q_0	3	Reference patch quality
m	2	Interference
N		Group size
N_P		Number of patches

stage is denoted $A_i = \{0, 1, 2, \dots, i-1, i+1, \dots, N\}$, where $j \in A_i$ is the rank of the recipient of the act of grooming, and where $j = 0$ means that individual i grooms no one. Grooming incurs a payoff cost c to the groomer, while contributing a payoff benefit b to the groomee (see Table 1 for a list of parameters).

Stage 2: Selecting a food patch

Each individual selects a food patch out of N_P available patches to feed on. We denote by $P = \{1, 2, \dots, N_P\}$ the set of food patches so that action $k \in P$ means an individual has selected food patch k . A food patch is defined as an area containing one or more sources of food, from which multiple individuals can feed, but that can also be monopolized by a single individual.

Stage 3: Aggressive interactions

After all individuals have chosen a patch, each makes a binary decision for every other individual on the same patch, whether to engage in a contest or not. Because we assume that individuals can only engage in a contest with others on the same food patch, we denote by $E_{i,t} \subseteq A_i$ the set of individuals that are on the same patch as i at time t .

After each individual has selected its targets (or no one) to attack, a contest occurs between the aggressor and the target, and the loser of this contest gets evicted from the patch. We adhere to the following rules for contests. (1) Individual i engaging in a contest with j is sufficient for a contest to occur between i and j (it is not required that j also wants to engage in a contest with i). (2) To each individual we associate an array of target individuals. The order in which the contests occur is determined by random sampling (without replacement) from this array. (3) Individuals that have lost a single contest will not engage in any further contests in the current time step, as they are evicted from the patch. This thus means that selecting a target to attack does not necessarily mean that the focal individual and the target will have a contest, as one (or both) of them may be evicted by another individual before the contest occurs.

We assume that the contest does not incur any costs, as these interactions are often settled without physical contact (covert interactions). The outcome of each interaction is given by a contest success function (e.g. Hirshleifer, 1989) so that individual i interacting with j wins the contest with probability

$$\frac{\exp(dr_i)}{\exp(dr_i) + \exp(dr_j)}, \quad (1)$$

where $d \geq 0$ is the dominance gradient, which determines the influence of the ranks of the individuals on the outcome of the contest. If $d = 0$, then rank has no influence on the outcome and both individuals are equally likely to win, but if $d \rightarrow \infty$ the higher-ranked individual is almost certain to win the contest.

Stage 4: Feeding

After aggression has occurred, individuals feed. We use the interference model to compute the payoff individuals gain on a given patch (Parker & Sutherland, 1986; Sutherland & Parker, 1992). The payoff per individual on patch k at time t is given by

$$\pi_{k,t}^P = \frac{Q_k}{N_{k,t}^m}, \quad (2)$$

where $Q_k > 0$ is the quality of patch k , $N_{k,t}$ is the number of individuals on patch k at time t during the feeding stage and thus after contests and eviction have occurred. The parameter $m \geq 0$ describes the interference between individuals on a food patch (a

larger m resulting in a greater decrease in payoff when an additional individual feeds on the patch). Thus, individuals benefit from evicting others as the payoff increases when there are fewer competitors. After all individuals have fed, the group mixes again so that in the following time step each individual can have a grooming interaction with any other individual of the group again. As we are not interested here in the survival effects of feeding, we assume that evicted individuals feed on a low-quality source that is abundant enough to ensure their survival.

We consider three different scenarios regarding how the quality of patches is determined. (1) All patches are of equal quality, so that $Q_k = q_0$ for all k and q_0 is used as a baseline quality. (2) Half the patches are of high quality (equal to q_0), while the other half are low quality (equal to $0.5q_0$). In cases where the number of patches is odd, there is one extra high-quality patch. (3) The quality of patch k is given by

$$Q_k = q_0 \left(\frac{1}{1.2} \right)^{(k-1)}, \quad (3)$$

thus, all patches differ in quality, where the quality asymptotically decreases with increasing k .

Learning and Decision-making Processes

Individuals take actions in three stages during an interaction step: whom to groom, select a food patch and whom to evict. We assume that individuals learn which action to play by way of reinforcement learning. We follow standard reinforcement learning models as used in biology and game theory (e.g. Achbany, Fous, Yen, Pirotte, & Saerens, 2006; Arbilly, Motro, Feldman, & Lotem, 2010, 2011a, 2011b; Camerer & Hua Ho, 1999; Dridi & Lehmann, 2014, 2015; Fudenberg & Levine, 1998; Hamblin & Giraldeau, 2009; Harley, 1981; McNamara & Houston, 1985), and assume that each individual has internal motivations (or attractions) for each action, which get updated as a function of time according to payoff received and allows the individual to express actions in a probabilistic way (e.g. Camerer & Hua Ho, 1999; Dridi & Lehmann, 2014, 2015; Fudenberg & Levine, 1998).

Action choice and motivation dynamics

Action choice for grooming (stage 1). The probability $p_{i,t}^G(j)$ that individual i takes action $j \in A_i$ during the grooming stage (i.e. grooms individual $j \neq i$ or grooming no one if $j = 0$) at time step t is given by

$$p_{i,t}^G(j) = \frac{\exp[\lambda M_{i,t}^G(j)]}{\sum_{q \in A_i} \exp[\lambda M_{i,t}^G(q)]}, \quad (4)$$

where $M_{i,t}^G(j)$ is the motivation that individual i has at time t for action $j \in A_i$ (and can be thought of as the cumulative reward up to t of taking that action) and is assumed to be a real-valued number ($M_{i,t}^G(j) \in \mathbb{R}$). Each individual takes a single action $j \in A_i$ each time step. Equation (4) is the standard logit choice rule for reinforcement learning of decision theory and neuroscience (e.g. Camerer & Hua Ho, 1999; Fudenberg & Levine, 1998; Sutton & Barto, 1998), and has been used in behavioural ecology (e.g. Arbilly et al., 2010, 2011a, 2011b; Dridi & Lehmann, 2015; see Dridi & Lehmann, 2014 for more details and justifications on the framework we use here). The parameter $\lambda \geq 0$ therein can be seen as the sensitivity to motivations or exploration rate: if $\lambda = 0$ each action is chosen with equal probability regardless of motivation; if $\lambda \rightarrow \infty$ the action with the highest motivation is chosen almost deterministically.

The change $\Delta M_{i,t}^G(j) = M_{i,t+1}^G(j) - M_{i,t}^G(j)$ in motivation (or attraction) of individual i for action $j \in A_i$ at time step $t \geq 1$ is assumed to take the form

$$\Delta M_{i,t}^G(j) = -\delta M_{i,t}^G(j) + \pi_{ij,t}^G + \pi_{ij,t}^F. \tag{5}$$

The first term on the right-hand side is a decay of motivation (for instance due to loss of memory) of i for action j , with discount factor δ . The second term, $\pi_{ij,t}^G$, reflects the change in motivation due to grooming and is given by

$$\pi_{ij,t}^G = \begin{cases} bl_{j \rightarrow i,t}^G - cl_{i \rightarrow j,t}^G, & \text{if } j \in A_i \setminus 0 \\ \gamma l_{i \rightarrow j,t}^G, & \text{if } j = 0, \end{cases} \tag{6}$$

where $l_{i \rightarrow j,t}^G$ is an indicator function equal to 1 if i grooms j at time t and 0 otherwise. The payoff γ that i obtains if it grooms no one is the reinforcement of not grooming again in the next time step. Assuming $b > 0$ and $c > 0$, the motivation of individual i to groom j thus increases with receiving grooming from j , but decreases if i grooms j .

The third term in equation (5) is the (motivational) payoff

$$\pi_{ij,t}^F = l_{ij,t}^C \pi_{k(ij),t}^P \left(\frac{1}{2} - \frac{\exp(dr_i)}{\exp(dr_i) + \exp(dr_j)} \right) \tag{7}$$

due to co-feeding. This depends on $l_{ij,t}^C$, which is equal to 1 if i and j feed on the same patch at time t and 0 otherwise. When i and j feed on the same patch, individual i obtains material payoff $\pi_{k(ij),t}^P$ (equation (2)), where $k(ij)$ is the patch on which i and j co-feed. We assume that the motivation reinforcement due to co-feeding (equation (7)) is proportional to this payoff and $(1/2 - \exp(dr_i)/[\exp(dr_i) + \exp(dr_j)])$. The term in parentheses causes the reinforcement to be positive if i has a lower rank than j , and negative if the rank of i is higher. Hence, we assume that a higher-ranked individual will decrease its motivation to groom a lower-ranked one after feeding together on a patch (since if it tolerates the lower-ranked individual, it should decrease its tendency to groom), while the low-ranked individual will increase its motivation to groom the high-ranked one. However, in the absence of a dominance gradient ($d = 0$), the reinforcement to grooming motivation is 0 regardless of rank.

We note that the reinforcement learning rule for motivation implemented by equation (5) (and those given below) is an example of the so-called relative payoff sum rule (Dridi & Lehmann, 2014; Hamblin & Giraldeau, 2009; Harley, 1981), where payoff over time is simply accumulated, and where the cumulated payoff decays at rate δ .

Action choice for patch selection (stage 2). The probability that at time t individual i selects patch $k \in P$ for feeding is assumed to be given by

$$p_{i,t}^P(k) = \frac{\exp[\lambda M_{i,t}^P(k)]}{\sum_{q \in P} \exp[\lambda M_{i,t}^P(q)]}, \tag{8}$$

which is similar to stage 1 (equation (4)) and where the change in the motivation $M_{i,t}^P(k)$ that individual i has at time t for patch $k \in P$ is updated as follows

$$\Delta M_{i,t}^P(k) = -\delta M_{i,t}^P(k) + \pi_{k,t}^P I_{ik,t}^P, \tag{9}$$

The first term is again the discounted motivation and the second term is the patch payoff to an individual multiplied by the indicator for the individual on the patch, which is equal to 1 if i feeds on patch

k at time t and 0 otherwise. Note that individuals that have been evicted from a patch do not positively reinforce their patch motivations.

Action choice for aggression (stage 3). To determine action choice for this stage, we let $M_{i,t}^T(j)$ represent the motivation of individual i to tolerate individual j (“the tolerance”).

The change $\Delta M_{i,t}^T(j) = M_{i,t+1}^T(j) - M_{i,t}^T(j)$ in tolerance is assumed to occur as follows:

$$\Delta M_{i,t}^T(j) = -\delta M_{i,t}^T(j) + bl_{j \rightarrow i,t}^G - \epsilon l_{ij,t}^A I_{ij,t}^E, \tag{10}$$

where the first term is the discounted motivation, the second term is the increase in tolerance if i has been groomed by j in round t (as in equation (6)), and the third term is the change in motivation as a result of the outcome of aggressive interactions, which depends on the parameter ϵ giving the impact of eviction on the motivation. In this third term, $I_{ij,t}^A$ takes value 1 if a contest occurred between i and j in round t and 0 otherwise, and $I_{ij,t}^E$ takes value 1 if i won the contest, and otherwise it takes value -1 (where the probability of winning a contest is given by equation (1)). Hence, the winner of the contest decreases its motivation to tolerate the loser of the contest, while the loser increases its motivation to tolerate the winner.

In the aggression stage, each individual takes a sequence of eviction decisions, one for each individual in its patch (one for each $j \in E_{i,t}$). We assume that the probability $p_{i,t}^T(j)$ that individual i wants to evict j is given by

$$p_{i,t}^T(j) = \frac{1}{1 + \exp[M_{i,t}^T(j) - \beta]}. \tag{11}$$

This is a logistic function bounded between 0 and 1 (see Fig. 1), where the probability of engaging in a contest decreases with increasing motivation for tolerance. As such, the function ensures that a higher motivation for tolerance for individual j results in a lower probability of engaging in a contest with that individual. The parameter β can be seen as the intrinsic aggressiveness of the individuals, as it regulates the probability that i will engage in a contest with j given its motivation to do so, since with the same motivation a larger value for β increases the probability that i engages in a contest with j (Fig. 1). According to equation (11) an individual makes a single decision for each other individual on the patch and can therefore attempt to evict anywhere between no one and all other individuals on the same patch.

We note that motivations can become negative in the model (the last term in equation (10) may induce negative motivation),

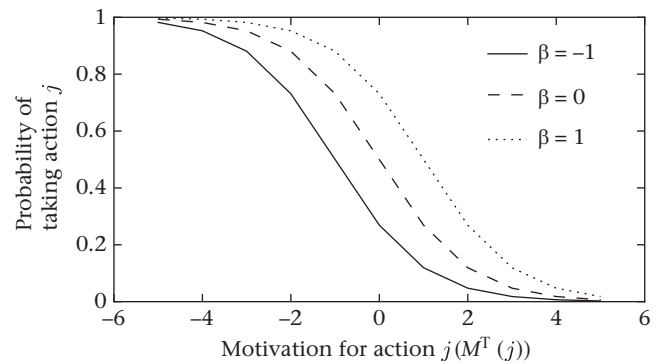


Figure 1. Probability of engaging in a contest with an individual as a function of motivation (equation (11)). Higher motivation for tolerance decreases the probability of engaging in a contest with the individual, while higher aggressiveness (β) increases the probability of engaging in a contest.

and so β can be negative. What matters for decision making is not the sign of the motivations, but only their ranking. The model could be rescaled such that motivations are only positive, but we refrain from doing so for simplicity of presentation and model analysis.

Descriptors of Interactions

To recapitulate, the model tracks the actions and the outcome of actions of each individual i for each round t of interaction, that is, whether or not it grooms individual j ($I_{i \rightarrow j,t}^G \in \{0, 1\}$), whether or not it feeds on patch k ($I_{ik,t}^P \in \{0, 1\}$), whether or not it tries to evict j ($I_{ij,t}^A \in \{0, 1\}$), whether it wins or loses a contest against j ($I_{ij,t}^E \in \{1, -1\}$), and thus to determine with whom it co-feeds ($I_{ij,t}^C \in \{0, 1\}$). This in turn will generate the networks of interactions, which will emerge from the decisions taken and are of two kinds: first, the grooming network (how much each individual grooms every other individual in the group), and second the feeding association network (proportion of time that pairs of individuals feed together), which can both be represented as graphs (Wasserman & Faust, 1994).

To describe these networks, we first let the learning process run for $T_{eq} < T$ time steps so that it reaches a quasi-equilibrium state and then characterize the networks in this state as follows. For the grooming network we computed the fraction g_{ij} of time that individual i spends grooming j at a quasi-equilibrium of the learning process. This is defined as

$$g_{ij} = \frac{1}{(T - T_{eq})} \sum_{t=T_{eq}+1}^T I_{i \rightarrow j,t}^G \tag{12}$$

where $T - T_{eq}$ is the amount of time we record the grooming interactions at quasi-equilibrium, whereby g_{i0} is the fraction of time i grooms no one. From g_{ij} , we also computed the average fraction of time that i receives grooming as

$$g_i = \frac{1}{N} \sum_{j=1}^N g_{ji} \tag{13}$$

whereby $1 - g_i$ gives the fraction of time i is not groomed.

Since we are interested in characterizing the grooming patterns, we compute the proportion of grooming actions that are directed up the hierarchy h_u , which varies between zero and one ($0 \leq h_u \leq 1$). If $h_u = 0.5$ the amount of grooming received by higher-ranked individuals is equal to the amount of grooming received by lower-ranked individuals, while if $h_u = 1$ then all grooming actions are directed towards an individual with a higher rank.

h_u is calculated as

$$h_u = \frac{1}{(T - T_{eq})} \sum_{t=T_{eq}+1}^T \frac{\sum_{i=1}^N \sum_{j>i}^N I_{i \rightarrow j,t}^G}{\sum_{i=1}^N \sum_{j=1}^N I_{i \rightarrow j,t}^G} \tag{14}$$

where the numerator in the second ratio is the number of grooming actions in the group at time t that are directed towards an individual with a higher rank, while the denominator is the total number of grooming actions at that time. Furthermore, to describe in more detail the pattern of grooming up the hierarchy, we also consider the fraction of grooming (among all grooming interactions) directed towards the alpha individual $h_\alpha \in [0,1]$. This is computed as

$$h_\alpha = \frac{1}{(T - T_{eq})} \sum_{t=T_{eq}+1}^T \frac{\sum_{i=1}^{N-1} I_{i \rightarrow N,t}^G}{\sum_{i=1}^{N-1} \sum_{j=1}^N I_{i \rightarrow j,t}^G} \tag{15}$$

where the numerator in the second ratio is the total number of grooming actions directed towards the highest-ranked individual (the ‘alpha individual’ which has rank N), and the denominator is the total number of grooming actions. The fraction of grooming that is directed one rank up the hierarchy ($h_1 \in [0,1]$) is computed as

$$h_1 = \frac{1}{(T - T_{eq})} \sum_{t=T_{eq}+1}^T \frac{\sum_{i=1}^{N-1} I_{i \rightarrow i+1,t}^G}{\sum_{i=1}^{N-1} \sum_{j=1}^N I_{i \rightarrow j,t}^G} \tag{16}$$

To characterize the feeding association network, we compute the frequency of feeding association between individuals i and j (a_{ij}) as,

$$a_{ij} = \frac{1}{(T - T_{eq})} \sum_{t=T_{eq}+1}^T I_{ij,t}^C \tag{17}$$

where $I_{ij,t}^C$ is the co-feeding indicator function. We also compute the (quasi-equilibrium) frequency that individual i is not evicted and thus feeds on a patch (f_i) as,

$$f_i = 1 - \frac{1}{(T - T_{eq})} \sum_{t=T_{eq}+1}^T \sum_{j=1}^N -I_{ij,t}^A 0.5 (I_{ij,t}^E - 1) \tag{18}$$

where the second term is the fraction of time individual i has been evicted. We subtract 1 from $I_{ij,t}^E$ to avoid counting the instances where individual i has won a contest (where $I_{ij,t}^E = 1$). Finally, we compute (at the quasi-stationary equilibrium), the average level of tolerance $\tau \in [0,1]$ over all individuals in the population, which is defined as the fraction of time that an individual has not engaged in a contest with a partner, given that it is not alone on its patch. If $\tau = 0$, an individual never tries to evict a partner, while if $\tau = 1$ an individual always try to evict a partner, given that it is not alone on a patch.

Because the model is stochastic and highly nonlinear we implemented it via individual-based simulations and tracked the actions and statistical descriptors numerically (simulation file is available on Figshare, <https://doi.org/10.6084/m9.figshare.5472811>). For all reported results (see next section) we set all motivations of all individuals initially to 0 (i.e. $M_{i,0}^G = M_{i,0}^P = M_{i,0}^T = 0$), which implies that actions are uniformly distributed, and then let the model run for $T = 5000$ time steps while using a period of $T_{eq} = 4000$ to reach the quasi-equilibrium state. We verified that $T_{eq} = 4000$ time steps were generally enough to reach a quasi-equilibrium state, as the pattern of variation in the hierarchy score (h_u) between replicates remains stable after 500–2000 time steps (see the Appendix). Assuming about seven grooming interactions per day (e.g. rhesus monkeys, *Macaca mulatta*, Lindburg, 1971), a period of 4000 time steps means a learning period of 1.5 years. To assess sensitivity to initial conditions we ran 10 replicates for each set of parameter values, and thus when we report hierarchy scores (h_u, h_α, h_1) and tolerance level (τ), we report their means (and sometimes standard errors).

The outcome of the model will be the emergence of different grooming and association networks, depending on the intensity of food competition and other model parameters. Although the model uses a large number of parameters, our main interest is in the parameters that regulate the intensity of the competition and the value of the rewards (Table 1). The main parameters varied

here are the number of patches (N_p), the aggressiveness (β), the benefit of grooming (b), the dominance gradient (d) and the effect of patch quality (q_0), while we investigate the effect of varying group size (N) and the level of interference (m) in the [Appendix](#).

RESULTS

To gain intuition about the outcomes of the model, we first consider two special cases: grooming without food competition (that is, no patch selection, aggressive interactions and feeding) and food competition without grooming. Unless stated otherwise, all results are based on the following baseline parameter values: $c = 1$, $b = 3$, $\gamma = 0.5$, $\lambda = 0.25$, $\delta = 0.1$, $\beta = 1$, $\epsilon = 1$, $d = 2$, $q_0 = 6$, $m = 0.5$, $N = 10$ ([Table 1](#)).

Uncoupled Model

Grooming for grooming

To understand the incentive structure in our model, we start with the simplest case of grooming interactions for group size $N = 2$ (skipping stages 2, 3 and 4 of an interaction round). In this case, grooming motivations are only reinforced by receiving grooming (equation (5) with $\pi_{ij,t}^F = 0$ always). This allows us to investigate the conditions under which grooming as an equilibrium behaviour emerges (as opposed to choosing the nongrooming action '0'). For this case, we show in the [Appendix](#) that when λ dominates δ ($\lambda \gg \delta$ so that reinforcement dominates stochastic choice), the change in the probabilities $p_{1,t}^G(2)$ and $p_{2,t}^G(1)$ that individual 1 grooms 2 and 2 grooms 1, respectively, can be approximated by the following differential equations.

$$\begin{aligned} \frac{dp_{1,t}^G(2)}{dt} &= \lambda \left(1 - p_{1,t}^G(2) \right) \left[p_{2,t}^G(1)b - \left(p_{1,t}^G(2)c + \left(1 - p_{1,t}^G(2) \right) \gamma \right) \right] \\ \frac{dp_{2,t}^G(1)}{dt} &= \lambda \left(1 - p_{2,t}^G(1) \right) \left[p_{1,t}^G(2)b - \left(p_{2,t}^G(1)c + \left(1 - p_{2,t}^G(1) \right) \gamma \right) \right]. \end{aligned} \quad (19)$$

The first term in brackets in each equation shows that an individual will increase the tendency to groom its partner proportionally to the level of grooming of the partner multiplied by the benefit b ; this captures reciprocity. The individual will decrease its grooming according to a term (second one in parentheses in each equation) that can be thought off as the average of the cost c of grooming and the reward γ of not grooming. Equation (19) makes explicit that if both interacting partners have an initial tendency to groom ($p_{1,0}^G(2) > 0$ and $p_{2,0}^G(1) > 0$), then it is possible that the learning dynamic converges to both individuals grooming each other. This is due to grooming being 'reciprocated' and thus reinforced over time. On the other hand, if an individual interacts with another that has a very low grooming tendency, then it will not get exploited and will reduce its level of grooming.

Equation (19) thus encapsulates the core behavioural assumptions behind our model and, from a quantitative point of view, the global dynamics of learning depends on the parameter values. For instance, while fixing $b = 3$ and $c = 1$, we find that for $\gamma = 0.5$ pairs always end up grooming each other as the reinforcement when choosing the nongrooming action is very low ([Fig. 2d](#)). For $\gamma = 2$, which is equal to $b - c$, the payoff of coordinating on either action will be the same to an individual, so we find that half the pairs end up grooming, while the other half ends up not grooming ([Fig. 2e](#)). For $\gamma = 5$ the reinforcement for not grooming is larger than

grooming an individual that grooms back ($\gamma > b - c$), and most pairs end up not grooming ([Fig. 2f](#)). While the emergence of grooming for grooming is affected by the differences between b and c , we note that the parameter λ , tuning the exploration of actions, also affects the outcomes, and thus needs to be rescaled to obtain the same qualitative results if b and c are varied proportionally. This is so because under reinforcement learning, exploration of actions depends on the absolute value of the payoffs and only when λ becomes very large (so that there is no exploration of action, equation (4)) would only the relative values of c and b matter for the outcome; that is, only the relative payoff of actions matters.

For the remainder of the results we set $\gamma = 0.5$, as we want grooming to emerge, and then determine when it will be directed up the hierarchy. For instance, for $N = 10$ with $\gamma = 0.5$ we find that (while still skipping the stages concerning food patch selection, aggressive interactions and feeding) there exists virtually no variation in grooming propensity between individuals, and all individuals spend at least 98% of their time grooming others. Although giving grooming decreases the motivation to repeat that action, giving grooming to an individual that gives grooming back results in an overall increase in the motivation to groom that individual (since we always assume $b > c$ in equation (6)). As such, we find (as expected) that each individual converges on predominantly grooming a single individual that also almost exclusively grooms back ([Fig. 3a](#)). Pairs of individuals thus trade equal amounts of grooming in the absence of food competition. This corresponds to the time matching of grooming that is regularly found in primates (e.g. [Schino, Ventura, & Troisi, 2003](#)).

Tolerance without grooming

If the grooming stage is skipped, individuals can still select a patch to feed, evict targets and feed, but there can be no grooming for tolerance. In this case, tolerance motivation no longer depends on grooming (equation (10) with $I_{j \rightarrow i,t}^G = 0$ always) but can still depend on the contests that have occurred, and the number of contests depends on the intrinsic aggressiveness of the individuals (equation (11)). We thus find that a larger β (higher aggressiveness) results in less tolerance ([Fig. 3b](#)). Additionally, the number of patches (N_p) influences the average number of competitors on a patch, which in turn influences the likelihood of engaging in a contest with at least one other individual. Thus, fewer patches means more competitors on the same patch and therefore less tolerance ([Fig. 3b](#)).

Full Model

Grooming up the hierarchy for tolerance

We now analyse how grooming for tolerance occurs in the full model (allowing all four interaction stages to occur). We find that if there is food competition, not all individuals manage to feed on a patch and several individuals get evicted. Regularly, no more than two individuals feed on the same patch ([Fig. 4a, c](#)). A third (lower-ranked) individual on the same patch would not be tolerated by at least one other individual and thus gets evicted. Hence, individuals tend to feed as dyads.

Although the distribution of the highest-ranked individuals over the available food patches differs between replicates of simulations, the highest-ranked individuals consistently segregate over the patches to avoid competition with one another ([Fig. 4a, c](#)). Between the individuals that feed together on a patch, the lower-ranked individual always grooms the higher-ranked one more than the higher-ranked individual grooms back and is the only individual

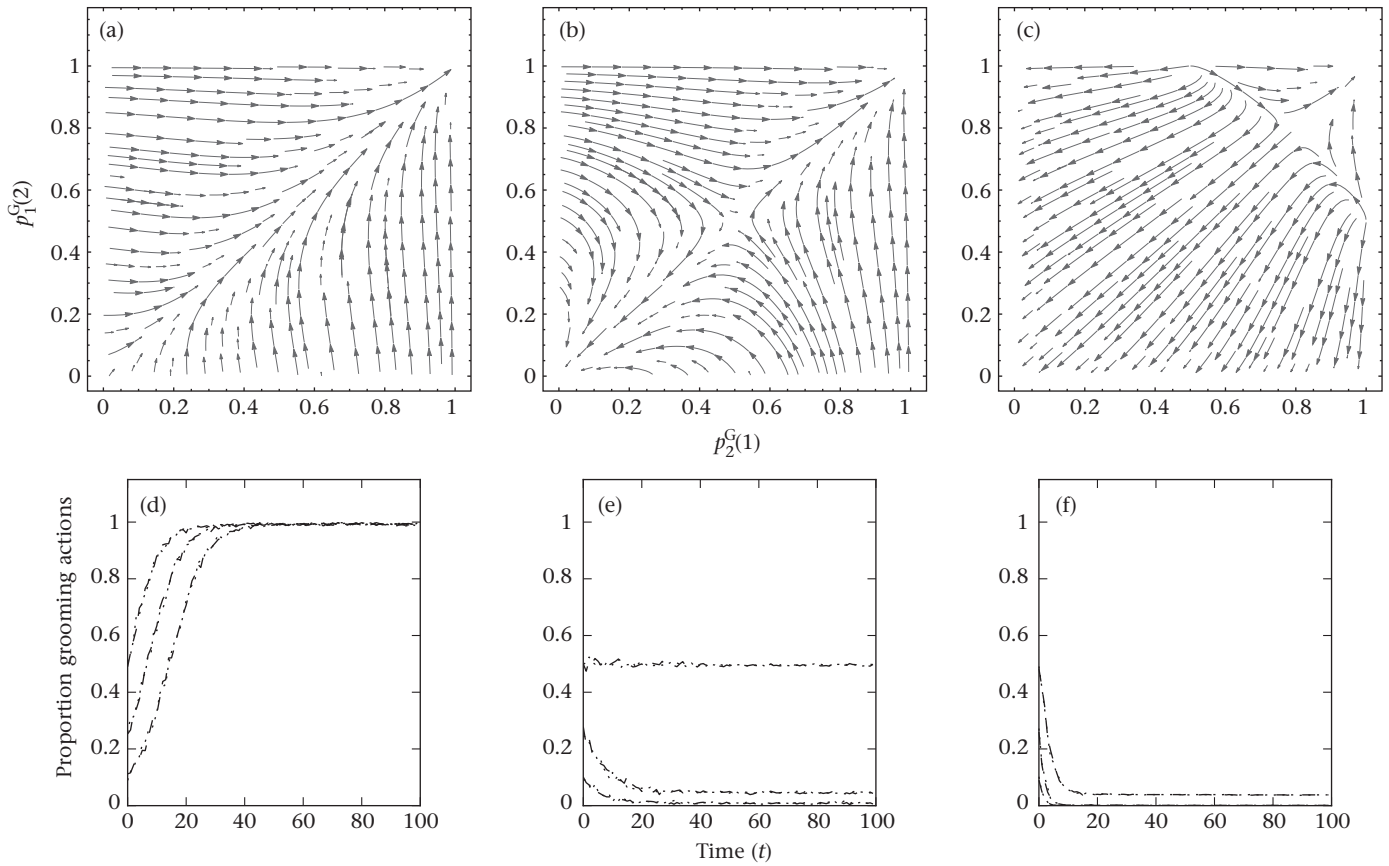


Figure 2. Solution orbits of the approximate learning dynamics given by equation (A1) in the Appendix (a refined version of equation (19) in the main text) and corresponding results obtained from individual-based simulations for the case: $c = 1$, $b = 3$, $\lambda = 0.25$, $\delta = 0.1$, $N = 2$. (a–c) The solution orbits describe the direction of change in the probability $p_1^G(2)$ ($p_2^G(1)$) that player 1 grooms 2 (2 grooms 1) and where the system will end up (either both partners groom each other or they do not groom). (d–f) Individual-based simulations showing the realizations of the (stochastic) model, averaged over 1000 replicates from three different initial conditions for the probability of grooming at $t = 1$: 0.5, 0.25 and 0.1 for every individual. (a, d) $\gamma = 0.5$, (b, e) $\gamma = 2$, (c, f) $\gamma = 5$. Note that for all panels $b - c = 2$. In (d–f) dashed lines are player 1 and dotted lines player 2 (the two lines strongly overlap, however, showing that the individuals converge simultaneously). Note that for $\gamma = 2$ a single player does not groom half the time, but instead half the pairs converge on grooming only and the other half on nongrooming only.

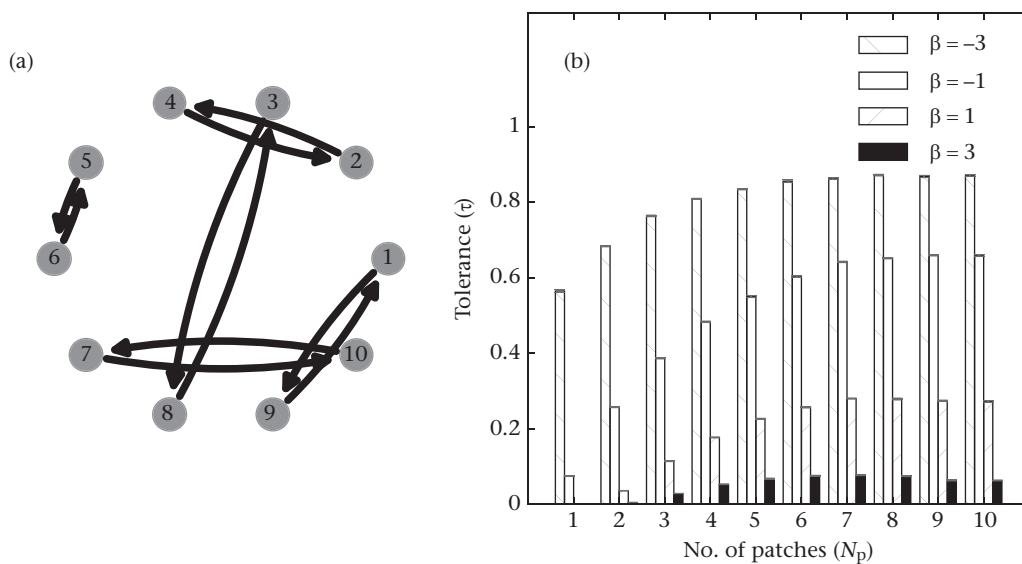


Figure 3. (a) Typical grooming network when only the grooming stage occurs (no patch selection, aggressive interactions or feeding). Vertex size correlates with the proportion of grooming i receives (g_i), while the edges correlate with the grooming i gives to j (g_{ij}). Individuals tend to form grooming pairs, where the members of the pair groom one another equally. For clarity, grooming proportions below 0.10 ($g_{ij} < 0.10$) are not plotted. (b) Tolerance (τ , mean and error bars) as a function of the number of patches, in the absence of grooming. Increasing aggressiveness (β) decreases the tolerance τ . Note that variation in tolerance is nearly absent, resulting in error bars drawn as a line at the top of each bar. Parameter values: $c = 1$, $b = 3$, $\gamma = 0.5$, $\lambda = 0.25$, $\delta = 0.1$, $e = 1$, $d = 2$, $q_0 = 6$, $m = 0.5$, $N = 10$, with patch quality given by equation (3).

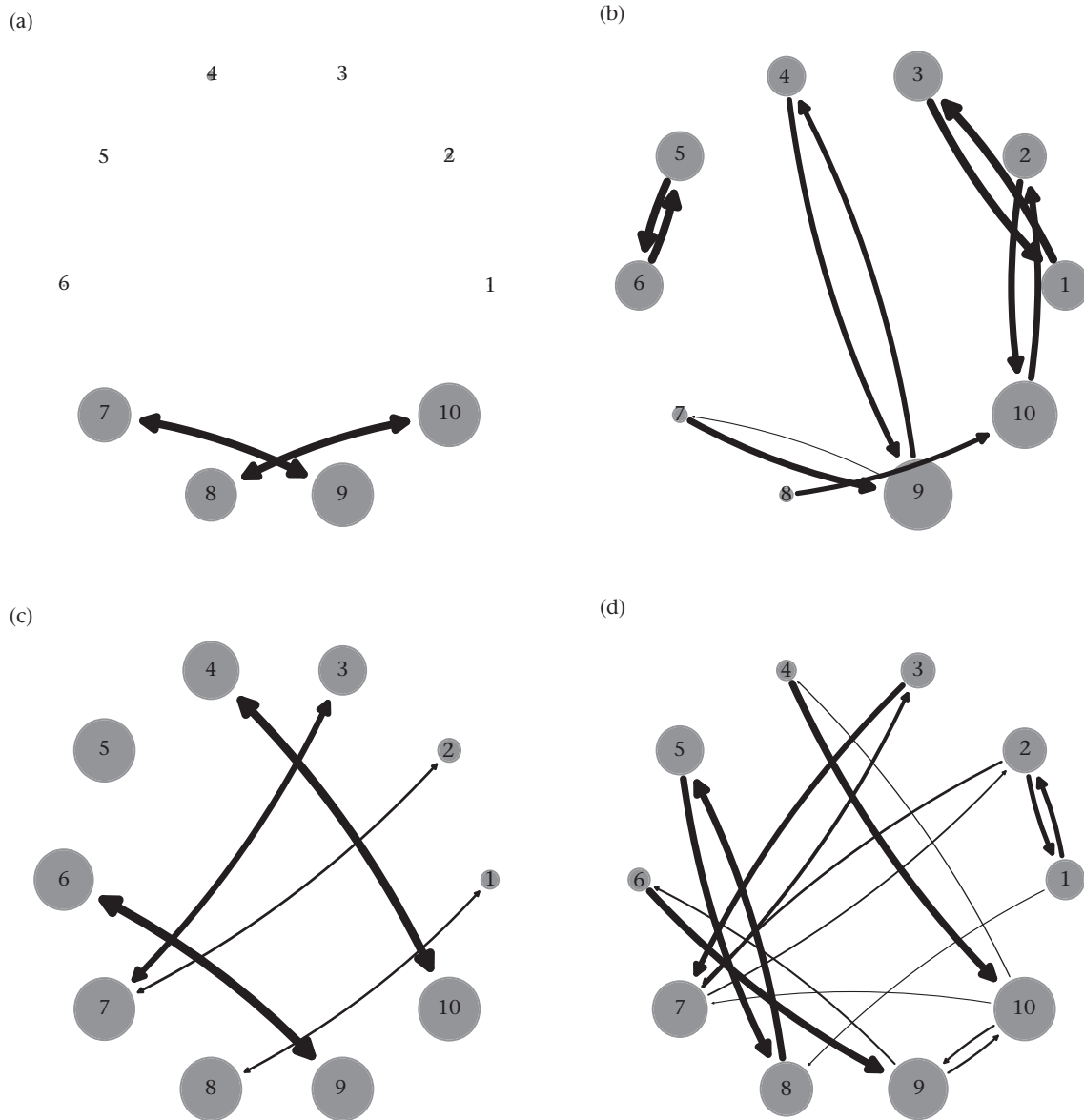


Figure 4. Typical (a, c) feeding association and (b, d) grooming networks for identical parameter values as in Fig. 3 and $\beta = 1$ but allowing for all interactions. (a, b) $N_p = 2$, (c, d) $N_p = 5$. In the feeding association network, the vertex size correlates with the frequency i feeds (f_i) while the edges correlate with the frequency i and j co-feed (a_{ij}). In the grooming network the vertex size correlates with the proportion of grooming i receives (g_i), while the edges correlate with the grooming i gives to j (g_{ij}). For clarity, associations and grooming proportions below 0.10 are not plotted (i.e. $a_{ij} < 0.10$, $g_{ij} < 0.10$). Note that the numbers relate to the position in the hierarchy where individual 10 is the highest-ranked individual. Individuals that feed together form an asymmetrical grooming relationship, while individuals that do not feed together form a symmetrical relationship (or they do not groom each other).

that is tolerated by the higher-ranked one (Fig. 4b, d). This is quantitatively captured by our hierarchy size score, h_u , which is systematically above 0.5 (Figs 5 and 6). For individuals that do not co-feed on a patch, the grooming between them is either balanced or absent. Thus, only when individuals feed together is an imbalance of grooming within that pair created where the higher-ranked individual receives additional grooming. Owing to the imbalance of grooming between individuals sharing a patch (the lower-ranked grooming more), we can thus conclude that our model generates grooming for tolerance; namely, grooming is exchanged for tolerance during feeding.

We also find that grooming up the hierarchy follows a systematic qualitative pattern in which the more there is grooming up the hierarchy, the more this is directed towards the alpha individual. We indeed find a strong positive correlation between h_u and h_α

(Pearson correlation: $r = 0.48$, $P < 0.01$; Fig. 5). Conversely, grooming just one rank up the hierarchy (h_1) is negatively correlated with h_u (Pearson correlation: $r = -0.15$, $P = 0.15$). From the feeding associations, as in Fig. 4, this is not surprising, since closely ranked individuals segregate over the available food patches and thus do not need to groom each other in order to be tolerated on a patch. Occasionally, individuals with adjacent ranks may share a patch, but it is clear from our results that grooming up the hierarchy is a result of the highest-ranked individuals (especially the alpha) receiving additional grooming and not because individuals groom one rank up the hierarchy.

Conditions favouring grooming for tolerance

We now delineate how grooming up the hierarchy depends on various parameters, such as patch number and quality (affecting

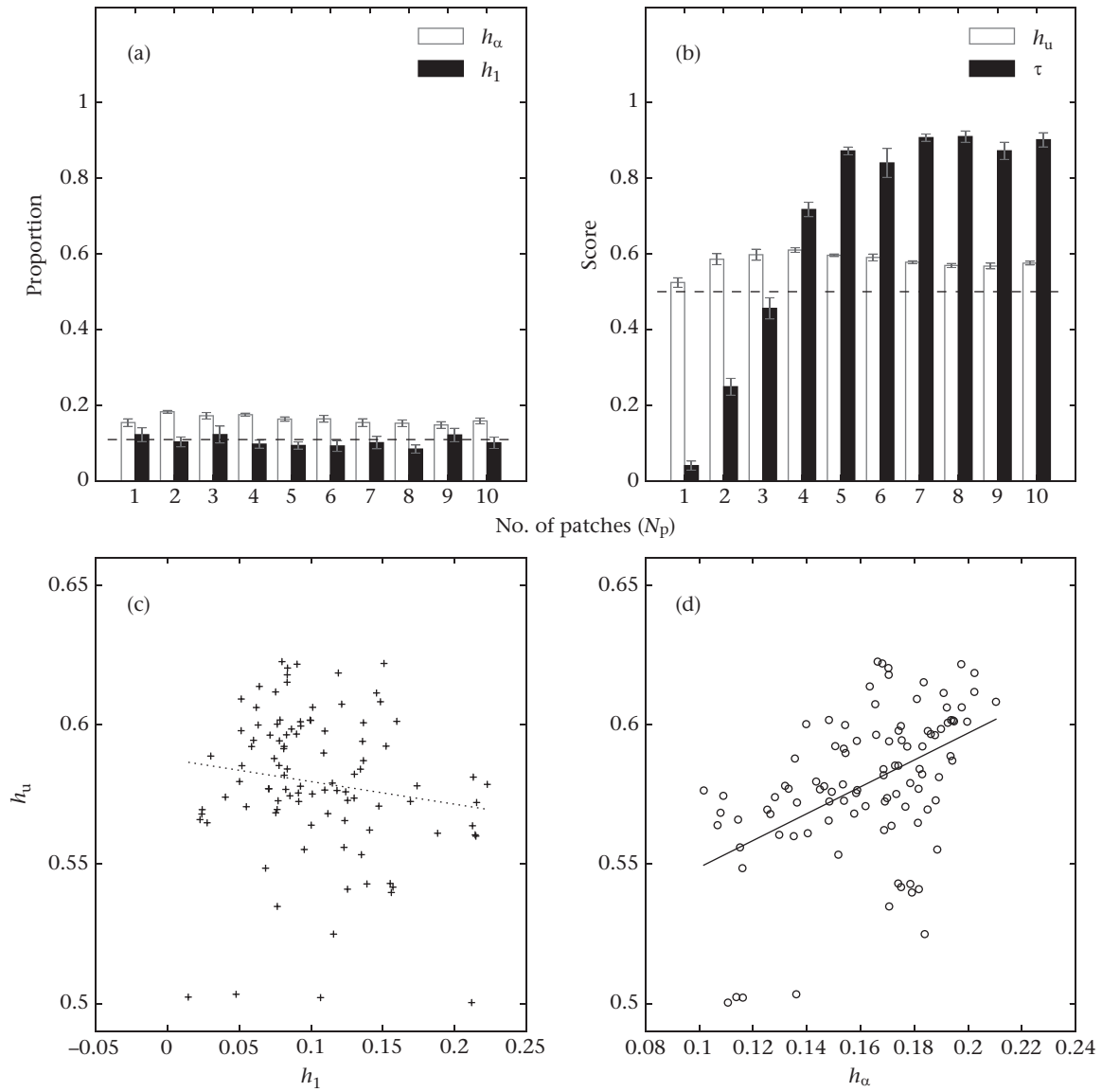


Figure 5. (a) Mean and standard error (over 10 replicates) of the proportion of grooming directed towards the alpha (h_α) and towards the individual one rank up the hierarchy (h_1) as a function of the number of patches. The dashed line shows the expected proportions if grooming actions were chosen randomly. (b) Mean and standard error of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches. The dashed line shows $h_u = 0.5$. (c) Scatter of h_u against h_1 , and the regression slope (dotted line). (d) Scatter of h_u against h_α , and the regression slope (solid line). Parameter values as in Fig. 4.

competition), aggressiveness and dominance gradient, and grooming costs and benefits.

We find that grooming up the hierarchy depends markedly on competition for resources. Indeed, the hierarchy score depends strongly on the number of available food patches (Fig. 5b). If there is just a single patch to feed on ($N_p = 1$), competition for feeding is high, and thus there is little tolerance ($\tau = 0.04$). Most individuals are evicted from the patch, and will not groom for tolerance, resulting in a hierarchy score close to 0.5 ($h_u = 0.52$). The slight amount of grooming that does occur up the hierarchy (since $h_u > 0.5$) is due to the second highest-ranked individual being tolerated and grooming the alpha.

If the number of food patches is increased, individuals will segregate over the available patches (see Fig. 4). Fewer individuals per patch means there will be less competition within each patch and thus more tolerance (Fig. 3b). The highest-ranked individual on

each patch is likely to tolerate one other individual in exchange for grooming (Fig. 4), and since these exchanges of grooming for tolerance occur more frequently if there are more patches, then there will consequently be more grooming up the hierarchy (Fig. 5b). Most grooming up the hierarchy occurs for $N_p = 4$ ($h_u = 0.61$). A further increase in the number of patches alleviates some of the competition, as individuals may sometimes find themselves alone on a patch, and therefore less tolerance and less grooming up the hierarchy are required. We thus observe that the hierarchy score is nonmonotonic in number of food patches, first increasing and then decreasing for $N_p > 4$ (Fig. 5b). However, even for high N_p there remains an overall tendency to groom up the hierarchy ($h_u = 0.58$ for $N_p = 10$), since the additional patches are of such low quality that individuals avoid feeding on these (quality of the best patch: $Q_1 = 6$; worst patch: $Q_{10} = 1.16$). In the Appendix we show that if there are as many patches as individuals and all

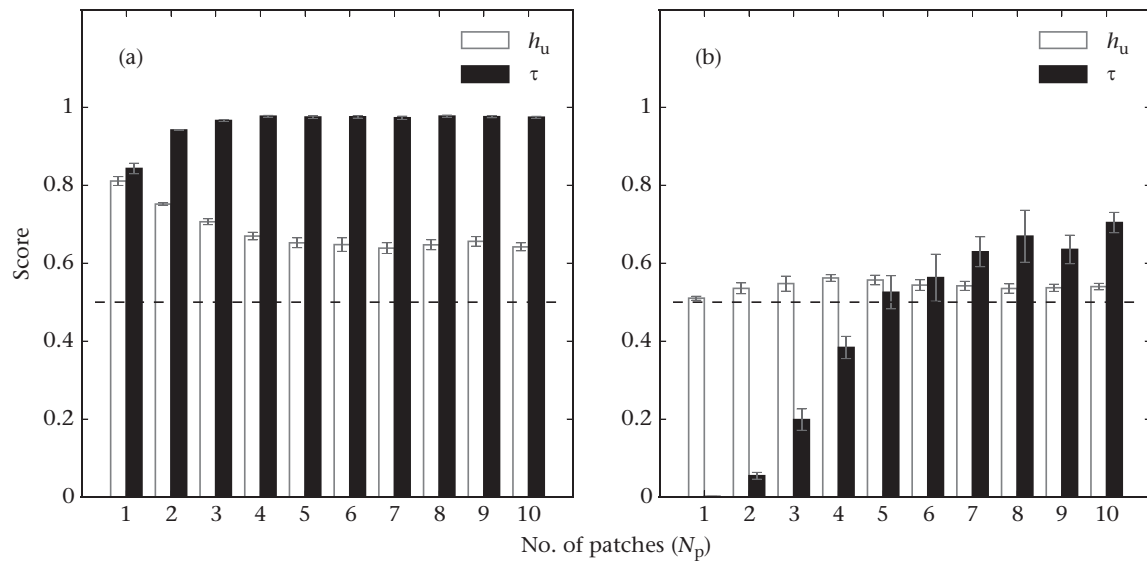


Figure 6. Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches. Parameter values as in Fig. 4. (a) $\beta = -3$, (b) $\beta = 3$. The dashed line shows $h_u = 0.5$.

patches are of equal quality, then grooming is completely reciprocal ($h_u = 0.50$), since there is no co-feeding and grooming for tolerance is not required (see section ‘Testing patch quality assumptions’ in the Appendix). Thus, in the absence of competition for resources we recover the grooming for grooming results (section ‘Grooming for grooming’).

For the effect of patch quality on grooming, we find that when this quality is ranked (equation (3)) lowering the baseline quality, q_0 , results in less grooming up the hierarchy. A lower patch quality means that individuals gain less material (and thus motivational) payoff on patches, and thus motivations for patch choice are less strongly affected by feeding (equation (9)). Individuals are then more likely to choose different patches at each time step and therefore regularly encounter different individuals on the patch. Strong feeding associations between individuals (i.e. feeding together regularly) will therefore not occur and thus there will be little grooming for tolerance (see Fig. A2a in the Appendix). Increasing patch quality increases grooming up the hierarchy, because higher-ranked individuals will be less inclined to groom lower-ranked individuals, while lower-ranked individuals will reward higher-ranking ones even more (equation (5)). Evicting another individual has a fixed effect on tolerance motivation, and the effect of being groomed outweighs the effect of evicting ($b > \varepsilon$). Therefore, if only the patch quality (q_0) changes, dominant individuals would continue to tolerate subordinates.

Grooming up the hierarchy also depends on the intrinsic aggressiveness of individuals (β). Lowering β means individuals are less likely to engage in contests with others to evict them (in stage 3 of the interaction stage), and individuals are tolerated more often (Fig. 3b). A consequence of more tolerance is that low-ranked individuals are more likely to reward higher-ranked individuals with grooming and grooming up the hierarchy increases (Fig. 6a). The effect of changing aggressiveness is most clear under conditions of high food competition (i.e. low number of food patches). For $N_p = 1$ and $\beta = -3$ the tolerance is greatly increased ($\tau = 0.84$) as well as the hierarchy score ($h_u = 0.81$). Conversely, higher aggressiveness ($\beta = 3$) increases the number of contests, thus lowering the tolerance. Then, if low-ranked individuals are not tolerated on food patches, they do not groom higher-ranked ones in return. As such,

increasing β results in less tolerance and thus grooming becoming more reciprocal (i.e. $h_u \rightarrow 0.5$), and for $N_p = 4$ we find $h_u = 0.56$ (Fig. 6b).

We find that a positive dominance gradient (d) is needed for grooming to be directed up the hierarchy. If $d = 0$, all individuals are symmetrical in the sense that each has a 0.5 probability to win a contest with every other individual (equation (1)). In this case, no individual can consistently control a food patch by evicting others. For $d = 0$ we thus find that $h_u = 0.5$ for any number of patches (N_p), showing that the intensity of food competition has no effect on grooming decisions. If individuals are symmetrical, grooming motivations are no longer reinforced due to co-feeding (equation (5) with $\pi_{ij,t}^f = 0$ always), and we thus find that grooming is exchanged for grooming (Fig. A3a). For $d = 2$ the probability of winning a contest with an individual one rank below is 0.88 (equation (1)) and thus a higher-ranked individual already wins most contests. We find that increasing the dominance gradient has little effect on the hierarchy scores as the contest success function (equation (1)) saturates (e.g. $d = 10$; Fig. A3b).

Finally, we investigate how the cost to benefit ratio of grooming affects grooming up the hierarchy by varying the benefit b (but always holding $b > c$). If b is large, grooming motivations are increasingly controlled by whom an individual is groomed (equation (5)), and thus grooming will become increasingly reciprocal (for $b = 4.5$ we find $h_u = 0.5$ regardless of number of patches; see Fig. A4b in the Appendix). The tolerance motivation similarly increases with b (equation (10)), and thus we continue to find individuals feeding together, but with reduced asymmetry of grooming between them. Hence, we have grooming for grooming with high tolerance. If b is low, we again observe a high tendency for reciprocal grooming, since tolerance motivations will be less affected by receiving grooming (see Fig. A4a in the Appendix). Concomitantly, one then has less tolerance overall. This shows that the benefit of grooming (b) has a nonmonotonic effect on grooming up the hierarchy.

Additional analyses and sensitivity of the results

In the Appendix we present several additional results to test the robustness of our qualitative results. We show that the

assumption regarding the distribution of patch quality (whether patches are of equal quality or of low and high quality, see equation (3)) has a surprisingly small effect on the outcome (see Fig. A5 in the Appendix). We also vary the group size N and the interference while feeding (m) and find that we can reproduce qualitatively similar levels of grooming up the hierarchy (see Figs A6 and A7 in the Appendix). Increasing group size lowers grooming up the hierarchy slightly, however, since more dyads groom reciprocally, while increasing interference has a similar effect as lowering patch quality (q_0) as both affect the payoff to an individual on a patch, and thus both influence the strength of feeding associations between individuals. Additionally, we investigate whether a cost of grooming ($c > 0$) is necessary for grooming to be directed up the hierarchy holding the benefit b fixed. Like the case where $c = 1$ (Fig. 5), we find that with $c = 0$ grooming is directed up the hierarchy, with most grooming up the hierarchy for an intermediate number of patches (Fig. A8). However, compared to the case where $c = 1$ (Fig. 5), there is less grooming up the hierarchy. This is because the absence of a cost favours grooming being traded for grooming (equation (19)) while it has no direct impact on the tolerance motivation (which is mainly affected by b , equation (10)). Overall, these results support the findings presented in the section ‘Conditions favouring grooming for tolerance’.

Furthermore, we attempt to create conditions where individuals preferentially groom one rank up the hierarchy (instead of the alpha individual). We do this by having patch payoff decrease dramatically if more than two individuals feed on the same patch. Then, by creating substantial differences in patch quality we expect that mostly the highest-ranked individuals would feed together on the highest-quality patch, the third- and fourth-ranked individuals on the second highest-quality patch, and so forth. However, we find even under such conditions that grooming up the hierarchy is preferentially directed towards the alpha individual ($h_x > h_1$, see Fig. A9 in the Appendix).

DISCUSSION

Grooming is maybe the most documented social behaviour in the primate literature, as it is relatively easy to quantify and understanding its occurrence is a major goal (Schino, 2001; Seyfarth, 1980). Grooming is regularly found to be directed up the hierarchy, where lower-ranked individuals groom higher-ranked ones. It has been argued that this occurs when there is competition for food, where low-ranked individuals trade grooming for either tolerance or coalitionary support from high-ranked ones. To better understand the conditions under which this is likely to occur, we here developed a quantitative reinforcement learning model, where individuals make decisions on whom to groom, where to feed, whom to tolerate on a food patch and where these actions depend dynamically on payoffs. With this model we have investigated under which conditions food competition can result in grooming being directed up the hierarchy (qualitatively), which grooming patterns emerge (which individuals receive additional grooming) and under which conditions it is likely to occur (and break down).

Grooming for Tolerance

Our main finding is that low-ranked individuals are indeed likely to trade grooming for tolerance from high-ranked individuals (Fig. 4) leading to grooming being directed up the hierarchy (Fig. 5). This matches the empirical observations made in various primate species where grooming is regularly found to be traded for

tolerance (Borgeaud & Bshary, 2015; Carne et al., 2011; Tiddi et al., 2011; Ventura et al., 2006; Xia et al., 2012).

Our model thus demonstrates that grooming can be traded for tolerance as a result of food competition. Seyfarth's (1977) simulation model on grooming in female monkeys was seminal in explicitly investigating grooming patterns. This model was relatively simple in the sense that individuals only made grooming decisions and were assumed to have an innate preference to groom the highest-ranking individuals. However, under conditions where high-ranked individuals cannot provide an alternative benefit to grooming (e.g. tolerance or coalitionary support) or are not willing to give tolerance because there is more to gain from eating alone, low-ranked individuals should adjust their grooming preferences accordingly. With a fixed preference, individuals would always groom up the hierarchy. The current work is thus a logical next step, since we allowed grooming preferences (motivation) to vary between individuals and over time. Our model successfully produced, depending on the state of the environment, both grooming for tolerance and grooming for grooming, the latter not being possible in Seyfarth's model (1977). Our results thus demonstrate that dynamic motivations (or preferences) that are updated through learning can capture a wider range of grooming patterns, and that grooming up the hierarchy depends crucially on ecological and behavioural conditions.

The Pattern of Grooming Up the Hierarchy

We find that grooming up the hierarchy is mostly the result of the alpha individual receiving additional grooming, but not the individual one rank up the hierarchy (Fig. 5). It has been argued that all individuals in a primate troop should be willing to preferentially groom the alpha individual (followed by the second highest-ranked individual, etc.), but that at the same time there is competition for grooming access, which means that not all individuals can groom their preferred target. Consequently, most individuals may end up grooming adjacent-ranked individuals instead (Seyfarth, 1977, 1980). However, the evidence for both competition for grooming access and more grooming between adjacent-ranked individuals is mixed (in favour of competition for grooming access: Fairbanks, 1980; Schino, 2001, against: Henzi et al., 2003; Silk, 1982; in favour of grooming adjacent ranked: Seyfarth, 1977; Seyfarth, 1980, against: O'Brien, 1993; Thierry, Gauthier, & Peighnot, 1990). Therefore, it is worth investigating which conditions lead to competition for grooming access or increased grooming of adjacent-ranked individuals.

In our model we do not find preferential grooming of individuals one rank above in the hierarchy. Interestingly, we find that adjacent-ranked individuals tend to avoid one another in order to be the highest-ranked individual on their chosen food patch (Fig. 4). In the absence of feeding associations between adjacent-ranked individuals, there is no incentive for the lower-ranked individual to groom the higher-ranked one for tolerance. Therefore, the factors causing individuals to groom one rank up or down the hierarchy need further investigation, but it seems unlikely that this is directly caused by food competition. An interesting/logical extension of the current model would be to introduce some form of competition for grooming partners to investigate whether such a restriction will result in individuals grooming adjacent-ranked individuals more frequently.

Conditions Where Grooming Up the Hierarchy Occurs and Breaks Down

We have quantified how the intensity of food competition (due to the number of food patches, aggressiveness, etc.) modulates the

grooming up the hierarchy. The so-called socioecological model (Sterck et al., 1997) tries to understand how food competition, predation risk and infanticide risk determine various aspects of primate societies (e.g. patterns of migration, dominance hierarchies and grooming behaviour; van Schaik, 1996; Sterck et al., 1997; Koenig, 2002; Kappeler & van Schaik, 2002). While the socioecological model predicts that food competition may cause grooming to be directed up the hierarchy, it does not give quantitative predictions regarding how the intensity of food competition will modulate the grooming patterns (it only distinguishes between types of competition, Sterck et al., 1997).

Using the socioecological model as a premise for our own analysis, we find that our model produces patterns of grooming up the hierarchy due to food competition similar to findings in the primate literature (Carne et al., 2011; Tiddi et al., 2011; Ventura et al., 2006; Xia et al., 2012). Although it is argued in these studies that grooming is traded for tolerance on food sources, the intensity of food competition is rarely investigated. However, our model shows that the relationship between food competition and grooming up the hierarchy is not necessarily linear (Fig. 5). We find that grooming up the hierarchy is maximized for an intermediate number of food patches as these conditions facilitate grooming being traded for tolerance between pairs of individuals (Fig. 4d). Fewer available food patches (high food competition) decreases the number of individuals that manage to feed on a food patch and thus fewer individuals trade grooming for tolerance. In contrast, higher numbers of food patches will result in individuals feeding alone and in the absence of competition these individuals thus also do not trade grooming for tolerance. Therefore, to better understand observed patterns of grooming behaviour in primates, it is crucial to quantify the intensity of food competition.

Additionally, we identified conditions under which grooming will be traded for grooming instead of tolerance. For example, the model predicts that when there is a sufficient number of high-quality patches available grooming will be reciprocal (see Fig. A5b in the Appendix). This matches the finding by Pazol and Cords (2005) that, despite the presence of a dominance hierarchy, blue monkeys, *Cercopithecus mitis*, actively avoided one another while feeding, and high- and low-ranking individuals therefore received equal amounts of grooming. Similarly, in baboon, *Papio cynocephalus*, troops where individuals did not compete for resources, a strong time-matching pattern of grooming was found (Barrett et al., 1999).

Our model also predicts that high aggressiveness can lead to less grooming up the hierarchy (Fig. 6b). This conforms to the finding of Leinfelder et al. (2001) that grooming is solely traded for grooming, despite the presence of monopolizable food sources, and this is attributed to high food-related aggression. Obviously, aggressiveness varies between individuals, groups and species (Fairbanks et al., 2004; Isbell, 1991). However, for the sake of simplicity we used a single parameter to set the aggressiveness of all individuals here. This showed that high aggression results in grooming being traded for grooming. Further work is required to investigate how grooming networks are affected if aggressiveness can vary between individuals.

Finally, there may exist other factors that are not captured in our model, which also can favour reciprocal grooming over grooming up the hierarchy. For example, in a troop of blue monkeys grooming was found to be traded reciprocally, but this appeared to be the result of intense between-group competition, which increases the importance of group cohesiveness leading to equal shares of grooming within dyads (Rowell et al., 1991). Since we investigated

grooming only as a result of within-group competition, such effects could not occur in our model.

Emotional Bookkeeping

Although primates reciprocate grooming and trade grooming for other commodities, the mechanistic processes underlying these decisions are difficult to unravel. One of the hypothesized explanations for how primates take actions in a social context is through emotional bookkeeping (Aureli & Schaffner, 2002; Schino & Aureli, 2009, 2017). Primates are thought to have partner-specific emotions that guide their decision-making process. Receiving benefits, in various currencies, are translated into single variables (emotions) in an individual towards those that provided the benefit. Through social interactions individuals will form differential social relationships. The emotions can thus be seen as a measure of the relationship quality and will determine the actions an individual takes in a social context. Evidence supporting emotional bookkeeping has been found in various species (Schino & Pellegrini, 2009; Schino, Di Giuseppe, & Visalberghi, 2009).

Our model can be interpreted as an application of emotional bookkeeping, where the motivations for taking actions (equations (5), (9) and (10)) can be seen as emotional states, since they reflect the history of interactions between the individuals. Different currencies are translated into a single motivation: both tolerance and receiving grooming affect the grooming motivation (equation (5)). Our model provides a quantitative example of how partner-specific motivations (or emotions) may be updated through interactions with social partners or decay over time. Different types of interactions (grooming, co-feeding, contests) can all affect a single motivation, which consequently will determine the actions an individual takes that are associated with that motivation. Our model shows that through such updating mechanisms stable grooming bonds can be formed between individuals, while at the same time it allows for interchange of currencies (grooming for tolerance) between individuals of different ranks (Fig. 4).

Assumptions of the Model

Our model is a proximate one that captures grooming and feeding actions in a group of interacting individuals, and we did not study the evolution of the learning rules here. Therefore, we have made a variety of assumptions for this model, especially regarding how motivations are updated as a result of payoff received.

We used a reinforcement learning model (instead of an evolutionary model) because both food sources and grooming partners fluctuate widely in a primate's life (Borgeaud, Sosa, Bshary, Sueur, & van de Waal, 2016), which makes it likely that social strategies are learned as it allows an individual to respond to the variation in social partners and the environments it will experience over its life span. We assumed that individual i 's motivation to groom individual j increases when i is groomed by j , while it decreases when i grooms j . This updating matches the logic of reciprocal strategies occurring under repeated interactions, the most well-known, perhaps, being tit-for-tat, which can be seen as a simple learning strategy (Dridi & Lehmann, 2014). We also chose parameter values such that grooming is favoured in the model (see Fig. 2). The assumptions of our model thus do not favour 'cheating' behaviour (choosing not to groom), since grooming acts are likely to be reciprocated. We were interested in patterns of grooming behaviour here, and not in how individuals can avoid being cheated.

Regarding tolerance, we assumed that winning a contest would increase the probability of engaging in a contest with the same individual in future time steps (thus becoming less tolerant). Since evicting another individual from a food patch is likely to increase the payoff of the evicting individual in nature, repeating such an action will increase its payoff similarly in future time steps, and this learning rule is thus likely to increase the payoff overall. Finally, we assumed that receiving grooming will increase the tolerance towards the groomer, which has been shown experimentally (Borgeaud & Bshary, 2015), and we assumed that tolerating a lower-ranked individual causes the higher-ranked one to be less likely to groom that individual, which is a logical consequence if grooming is to be traded for tolerance only instead of both tolerance and grooming.

Although we made only assumptions regarding direct effects of actions and payoffs on motivations, these assumptions imply that various behavioural feedbacks will occur in the model. For example, while grooming motivations are not directly affected by aggressive interactions, these interactions do determine which individuals feed together on a patch, and co-feeding in turn does affect grooming motivations. As such, aggression feeds back on grooming decisions. Likewise, being evicted from a patch does not directly decrease the motivation to visit that patch in future time steps, but as the individual will no longer feed from it, over time it will learn not to visit the patch any more as a result of behavioural feedbacks. Hence, there is no need to consider any direct effects of aggression and eviction on motivations, and the model takes the consequences of these behaviours into account on action choice in a parsimonious way.

Outlook

Our model generates patterns of grooming reciprocity (trading grooming for grooming) and grooming up the hierarchy (trading grooming for tolerance), which matches a number of grooming patterns in natural populations. These different patterns emerge from the model when its parameters are varied. Therefore, it would be useful to test how well these parameters can explain the variation in grooming patterns in natural populations. For example, we have shown that the amount of grooming up the hierarchy does not necessarily increase with increasing food competition (Fig. 5), while high aggressiveness suppresses grooming up the hierarchy (Fig. 6). Using the wealth of primate data in the literature, such relationships can be tested quantitatively (e.g. using meta-analysis; Schino, 2001; Schino, 2007; Schino & Aureli, 2008). For example, the degree of reciprocity in grooming as a function of hierarchy can be evaluated per species (or population) in a similar way as h_u in our model. This index can then be correlated with a variety of other measures such as the frequency with which individuals feed together, the tolerance around food or the hierarchy steepness. While a good index for the degree of tolerance around food is difficult to quantify in natural populations, one could measure the maximum distance at which a dominant individual tolerates a subordinate with a certain probability (Borgeaud & Bshary, 2015). Alternatively, the number of individuals within a certain range (e.g. 5 m) during foraging could be used (as such data may be more readily available). Our model has yielded various predictions regarding how grooming up the hierarchy correlates with the number of food patches, aggressiveness, dominance gradient, etc. These predictions can thus usefully guide comparative studies and meta-analysis investigating grooming patterns and social relationships in primates.

Acknowledgments

We thank two referees for constructive comments. This work was supported by a grant from the Swiss National Science Foundation (grant number PDFMP3_137165).

References

- Achbany, Y., Fouss, F., Yen, L., Pirotte, A., & Saerens, M. (2006). Optimal tuning of continual online exploration in reinforcement learning. In S. Kollias, A. Stafylopatis, W. Duch, & E. Oja (Eds.), *Artificial Neural Networks—ICANN 2006, volume 4131 of lecture notes in computer science* (pp. 790–800). Berlin, Germany: Springer.
- Arbilly, M., Motro, U., Feldman, M. W., & Lotem, A. (2010). Co-evolution of learning complexity and social foraging strategies. *Journal of Theoretical Biology*, *267*(4), 573–581.
- Arbilly, M., Motro, U., Feldman, M. W., & Lotem, A. (2011a). Evolution of social learning when high expected payoffs are associated with high risk of failure. *Journal of the Royal Society Interface*, *8*(64), 1604–1615.
- Arbilly, M., Motro, U., Feldman, M. W., & Lotem, A. (2011b). Recombination and the evolution of coordinated phenotypic expression in a frequency-dependent game. *Theoretical Population Biology*, *80*(4), 244–255.
- Aureli, F., & Schaffner, C. (2002). Relationship assessment through emotional mediation. *Behaviour*, *139*(2), 393–420.
- Barrett, L., & Henzi, P. (2006). Monkeys, markets and minds: Biological markets and primate sociality. In P. Kappeler, & C. van Schaik (Eds.), *Cooperation in primates and humans* (pp. 209–232). Berlin, Germany: Springer.
- Barrett, L., Henzi, S. P., Weingrill, T., Lycett, J. E., & Hill, R. A. (1999). Market forces predict grooming reciprocity in female baboons. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1420), 665–670.
- Borgeaud, C., & Bshary, R. (2015). Wild vervet monkeys trade tolerance and specific coalitional support for grooming in experimentally induced conflicts. *Current Biology*, *25*(22), 3011–3016.
- Borgeaud, C., Sosa, S., Bshary, R., Sueur, C., & van de Waal, E. (2016). Intergroup variation of social relationships in wild vervet monkeys: A dynamic network approach. *Frontiers in Psychology*, *7*, 915.
- Camerer, C., & Hua Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, *67*(4), 827–874.
- Carne, C., Wiper, S., & Semple, S. (2011). Reciprocity and interchange of grooming, agonistic support, feeding tolerance, and aggression in semi-free-ranging Barbary macaques. *American Journal of Primatology*, *73*(11), 1127–1133.
- Cheney, D. L., Moscovice, L. R., Heesen, M., Mundry, R., & Seyfarth, R. M. (2010). Contingent cooperation between wild female baboons. *Proceedings of the National Academy of Sciences*, *107*(21), 9562–9566.
- Dridi, S., & Lehmann, L. (2014). On learning dynamics underlying the evolution of learning rules. *Theoretical Population Biology*, *91*, 20–36.
- Dridi, S., & Lehmann, L. (2015). A model for the evolution of reinforcement learning in fluctuating games. *Animal Behaviour*, *104*, 1–28.
- Dunbar, R. I. M. (1991). Functional significance of social grooming in primates. *Folia Primatologica*, *57*(3), 121–131.
- Dunbar, R. I. M. (2002). Modelling primate behavioral ecology. *International Journal of Primatology*, *23*(4), 785–819.
- Fairbanks, L. A. (1980). Relationships among adult females in captive vervet monkeys: Testing a model of rank-related attractiveness. *Animal Behaviour*, *28*(3), 853–859.
- Fairbanks, L. A., Newman, T. K., Bailey, J. N., Jorgensen, M. J., Breidenthal, S. E., Ophoff, R. A., et al. (2004). Genetic contributions to social impulsivity and aggressiveness in vervet monkeys. *Biological Psychiatry*, *55*(6), 642–647.
- Fudenberg, D., & Levine, D. K. (1998). *The theory of learning in games*. Cambridge, MA: MIT Press.
- Hamblin, S., & Giraldeau, L.-A. (2009). Finding the evolutionarily stable learning rule for frequency-dependent foraging. *Animal Behaviour*, *78*(6), 1343–1350.
- Harley, C. B. (1981). Learning the evolutionarily stable strategy. *Journal of Theoretical Biology*, *89*(4), 611–633.
- Henzi, P., & Barrett, L. (1999). The value of grooming to female primates. *Primates*, *40*(1), 47–59.
- Henzi, S. P., Barrett, L., Gaynor, D., Greeff, J., Weingrill, T., & Hill, R. A. (2003). Effect of resource competition on the long-term allocation of grooming by female baboons: Evaluating Seyfarth's model. *Animal Behaviour*, *66*(5), 931–938.
- Hirshleifer, J. (1989). Conflict and rent-seeking success functions: Ratio vs. difference models of relative success. *Public Choice*, *63*(2), 101–112.
- Isbell, L. A. (1991). Contest and scramble competition: Patterns of female aggression and ranging behavior among primates. *Behavioral Ecology*, *2*(2), 143–155.
- Kappeler, P. M., & van Schaik, C. P. (2002). Evolution of primate social systems. *International Journal of Primatology*, *23*(4), 707–740.
- Koenig, A. (2002). Competition for resources and its behavioral consequences among female primates. *International Journal of Primatology*, *23*(4), 759–783.
- Leinfelder, I., de Vries, H., Deleu, R., & Nelissen, M. (2001). Rank and grooming reciprocity among females in a mixed-sex group of captive hamadryas baboons. *American Journal of Primatology*, *55*(1), 25–42.

- Lindburg. (1971). *The Rhesus monkey in North India: An ecological and behavioral study*. New York, NY: Academic Press.
- Maestriperieri, D. (1993). Vigilance costs of allogrooming in macaque mothers. *American Naturalist*, 141(5), 744–753.
- McNamara, J. M., & Houston, A. I. (1985). Optimal foraging and learning. *Journal of Theoretical Biology*, 117(2), 231–249.
- O'Brien, T. G. (1993). Allogrooming behaviour among adult female wedge-capped capuchin monkeys. *Animal Behaviour*, 46(3), 499–510.
- Parker, G. A., & Sutherland, W. J. (1986). Ideal free distributions when individuals differ in competitive ability: Phenotype-limited ideal free models. *Animal Behaviour*, 34(4), 1222–1242.
- Pazol, K., & Cords, M. (2005). Seasonal variation in feeding behavior, competition and female social relationships in a forest dwelling guenon, the blue monkey (*Cercopithecus mitis stuhlmanni*), in the kakamega forest, Kenya. *Behavioral Ecology and Sociobiology*, 58(6), 566–577.
- Pusey, A. E., & Packer, C. (1987). Dispersal and philopatry. In B. B. Smuts, D. L. Cheney, R. M. Seyfarth, R. W. Wrangham, & T. T. Struhsaker Oja (Eds.), *Primate societies* (pp. 250–266). Chicago IL: University of Chicago Press.
- Roubová, V., Konečná, M., Šmilauer, P., & Wallner, B. (2015). Whom to groom and for what? Patterns of grooming in female barbary macaques (*Macaca sylvanus*). *PLoS ONE*, 10(2). e0117298+.
- Rowell, T. E., Wilson, C., & Cords, M. (1991). Reciprocity and partner preference in grooming of female blue monkeys. *International Journal of Primatology*, 12(4), 319–336.
- Sambrook, T. D., Whiten, A., & Strum, S. C. (1995). Priority of access and grooming patterns of females in a large and a small group of olive baboons. *Animal Behaviour*, 50(6), 1667–1682.
- van Schaik, C. P. (1989). The ecology of social relationships amongst female primates. In V. Standen, & R. A. Foley (Eds.), *Comparative socioecology. The behavioural ecology of humans and other mammals* (pp. 195–218). Oxford, U.K.: Blackwell Scientific.
- van Schaik, C. P. (1996). Social evolution in primates: The role of ecological factors and male behaviour. *Proceedings of the British Academy*, 88, 9–31.
- Schino, G. (2001). Grooming, competition and social rank among female primates: A meta-analysis. *Animal Behaviour*, 62(2), 265–271.
- Schino, G. (2007). Grooming and agonistic support: A meta-analysis of primate reciprocal altruism. *Behavioral Ecology*, 18(1), 115–120.
- Schino, G., & Aureli, F. (2008). Grooming reciprocation among female primates: A meta-analysis. *Biology Letters*, 4(1), 9–11.
- Schino, G., & Aureli, F. (2009). Reciprocal altruism in primates: Partner choice, cognition, and emotions. *Advances in the Study of Behaviour*, 39, 45–69.
- Schino, G., & Aureli, F. (2017). Reciprocity in group-living animals: Partner control versus partner choice. *Biological Reviews*, 92, 665–672.
- Schino, G., Di Giuseppe, F., & Visalberghi, E. (2009). The time frame of partner choice in the grooming reciprocation of *Cebus apella*. *Ethology*, 115(1), 70–76.
- Schino, G., & Pellegrini, B. (2009). Grooming in mandrills and the time frame of reciprocal partner choice. *American Journal of Primatology*, 71(10), 884–888.
- Schino, G., Ventura, R., & Troisi, A. (2003). Grooming among female Japanese macaques: Distinguishing between reciprocation and interchange. *Behavioral Ecology*, 14(6), 887–891.
- Seyfarth, R. M. (1977). A model of social grooming among adult female monkeys. *Journal of Theoretical Biology*, 65(4), 671–698.
- Seyfarth, R. M. (1980). The distribution of grooming and related behaviours among adult female vervet monkeys. *Animal Behaviour*, 28(3), 798–813.
- Seyfarth, R. M., & Cheney, D. L. (1984). Grooming, alliances and reciprocal altruism in vervet monkeys. *Nature*, 308(5959), 541–543.
- Silk, J. B. (1982). Altruism among female *Macaca radiata*: Explanations and analysis of patterns of grooming and coalition formation. *Behaviour*, 79(2), 162–188.
- Sterck, E. H. M., Watts, D. P., & van Schaik, C. P. (1997). The evolution of female social relationships in nonhuman primates. *Behavioral Ecology and Sociobiology*, 41(5), 291–309.
- Sutherland, W. J., & Parker, G. A. (1992). The relationship between continuous input and interference models of ideal free distributions with unequal competitors. *Animal Behaviour*, 44, 345–355.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
- Tanaka, I., & Takefushi, H. (1993). Elimination of external parasites(lice) is the primary function of grooming in free-ranging Japanese macaques. *Anthropological Science*, 101(2), 187–193.
- Thierry, B., Gauthier, C., & Peignot, P. (1990). Social grooming in tonkean macaques (*Macaca tonkeana*). *International Journal of Primatology*, 11(4), 357–375.
- Tiddi, B., Aureli, F., Polizzi di Sorrentino, E., Janson, C. H., & Schino, G. (2011). Grooming for tolerance? Two mechanisms of exchange in wild tufted capuchin monkeys. *Behavioral Ecology*, 22(3), 663–669.
- Ventura, R., Majolo, B., Koyama, N. F., Hardie, S., & Schino, G. (2006). Reciprocation and interchange in wild Japanese macaques: Grooming, cofeeding, and agonistic support. *American Journal of Primatology*, 68(12), 1138–1149.
- Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge, MA: Cambridge University Press.
- Wei, W., Qi, X.-G., Guo, S.-T., Zhao, D.-P., Zhang, P., Huang, K., et al. (2012). Market powers predict reciprocal grooming in golden snub-nosed monkeys (*Rhinopithecus roxellana*). *PLoS ONE*, 7(5). e36802+.
- Wrangham, R. W. (1980). An ecological model of female-bonded primate groups. *Behaviour*, 75(3), 262–300.
- Xia, D., Li, J., Garber, P. A., Sun, L., Zhu, Y., & Sun, B. (2012). Grooming reciprocity in female Tibetan macaques *Macaca thibetana*. *American Journal of Primatology*, 74(6), 569–579.

Appendix

Stochastic Approximation of Learning Dynamics

For the case $N = 2$ and assuming only grooming interactions (skipping stages 2, 3 and 4 of an interaction round), we can straightforwardly derive differential equations for the probabilities $p_{1,t}^G(2)$ and $p_{2,t}^G(1)$ that individual 1 grooms 2 and 2 grooms 1, respectively, by way of applying stochastic approximation as detailed in Dridi and Lehmann (2014, 2015). Indeed, using equations (5–6) of the main text (by setting $\pi_{ij,t}^F$), we have by using equations (11) and (13) of Dridi and Lehmann (2014) that

$$\begin{aligned} \frac{dp_{1,t}^G(2)}{dt} &= (1 - p_{1,t}^G(2)) \left(\delta \log \left(\frac{1 + p_{1,t}^G(2)}{p_{1,t}^G(2)} \right) + \lambda [p_{2,t}^G(1)b - (p_{1,t}^G(2)c + (1 - p_{1,t}^G(2))\gamma)] \right) \\ \frac{dp_{2,t}^G(1)}{dt} &= (1 - p_{2,t}^G(1)) \left(\delta \log \left(\frac{1 + p_{2,t}^G(1)}{p_{2,t}^G(1)} \right) + \lambda [p_{1,t}^G(2)b - (p_{2,t}^G(1)c + (1 - p_{2,t}^G(1))\gamma)] \right). \end{aligned} \quad (A1)$$

If we further assume that λ dominates δ ($\lambda \gg \delta$), then we can neglect the first terms in these equations (which account for randomness in choice) and simply write

$$\begin{aligned} \frac{dp_{1,t}^G(2)}{dt} &\sim \lambda (1 - p_{1,t}^G(2)) [p_{2,t}^G(1)b - (p_{1,t}^G(2)c + (1 - p_{1,t}^G(2))\gamma)] \\ \frac{dp_{2,t}^G(1)}{dt} &\sim \lambda (1 - p_{2,t}^G(1)) [p_{1,t}^G(2)b - (p_{2,t}^G(1)c + (1 - p_{2,t}^G(1))\gamma)], \end{aligned} \quad (A2)$$

which is the equation presented in the main text. To compare this with individual-based simulations (see Fig. 2), we used equation (A1).

Quasi-equilibrium

To determine whether $T_{eq} = 4000$ is a sufficient time period to reach the quasi-equilibrium, we investigated how the standard deviation of our main statistic, the hierarchy score (h_u), varied over time. We find in general that the standard deviation slowly drops in roughly the first 500–2000 time steps, after which it continues to vary, but the pattern of variation remains stable (Fig. A1). Therefore, a period of $T_{eq} = 4000$ interaction steps is sufficient to reach the quasi-equilibrium.

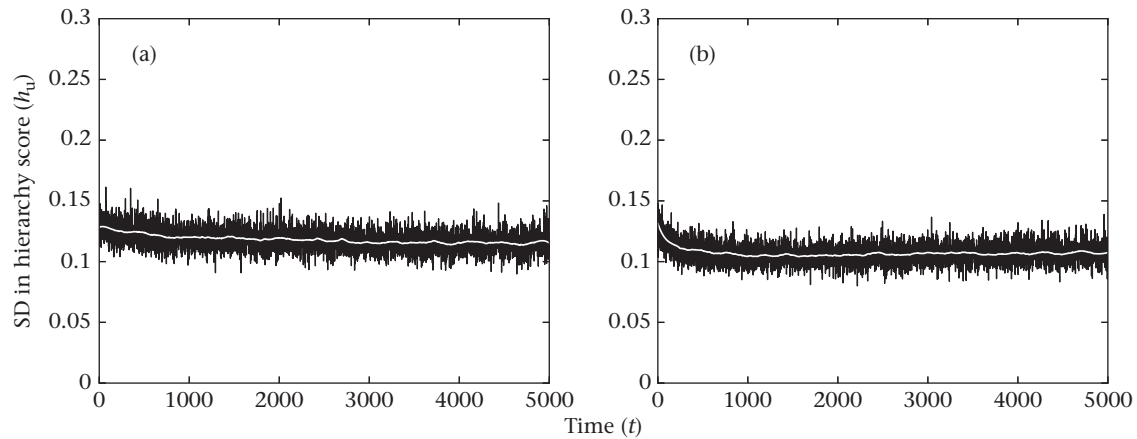


Figure A1. The standard deviation (over 100 replicates) in hierarchy score (h_u) as a function of time. Parameter values: $c = 1, b = 3, \gamma = 0.5, \lambda = 0.25, \delta = 0.1, \beta = 1, e = 1, d = 2, q_0 = 6, m = 0.5, N = 10$. (a) $N_p = 2$, (b) $N_p = 5$.

Conditions Where Grooming Up the Hierarchy Occurs and Breaks Down

Effect of patch quality

Fig. A2 illustrates the effect of varying the patch quality (q_0) on grooming up the hierarchy (h_u). The grooming up the hierarchy score increases with increasing patch quality (q_0).

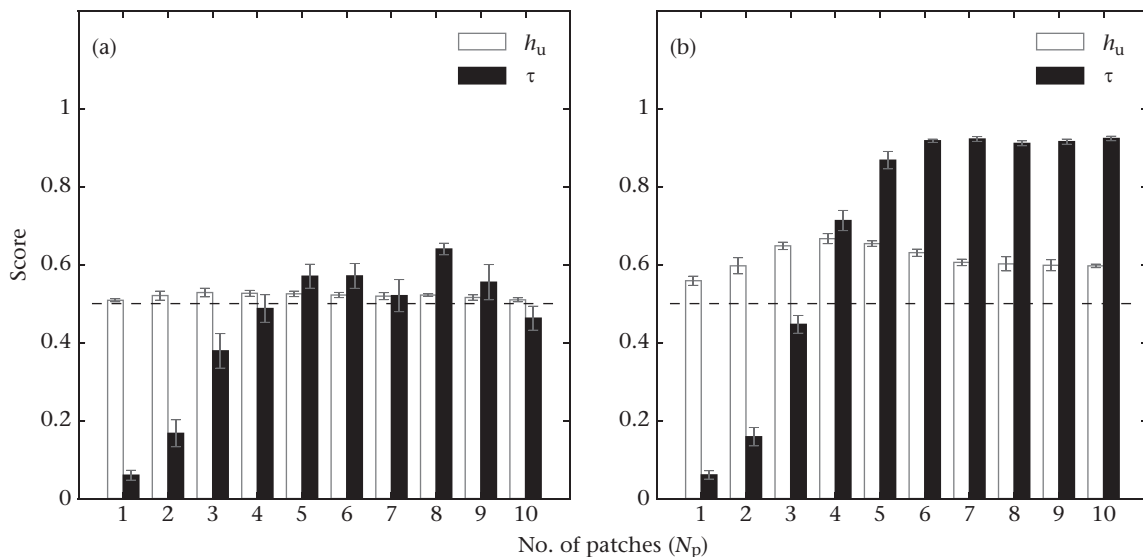


Figure A2. Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches, for low- and high-quality patches. Parameter values: $c = 1, b = 3, \gamma = 0.5, \lambda = 0.25, \delta = 0.1, \beta = 1, e = 1, d = 2, m = 0.5, N = 10$, with patch quality given by equation (3). (a) $q_0 = 3$, (b) $q_0 = 9$. The dashed line shows $h_u = 0.5$.

Effect of dominance gradient

Fig. A3 illustrates the effect of varying the dominance gradient (d) on grooming up the hierarchy (h_u). In the absence of a dominance gradient ($d = 0$, Fig. A3a) grooming is completely reciprocal ($h_u = 0.5$ for any number of patches).

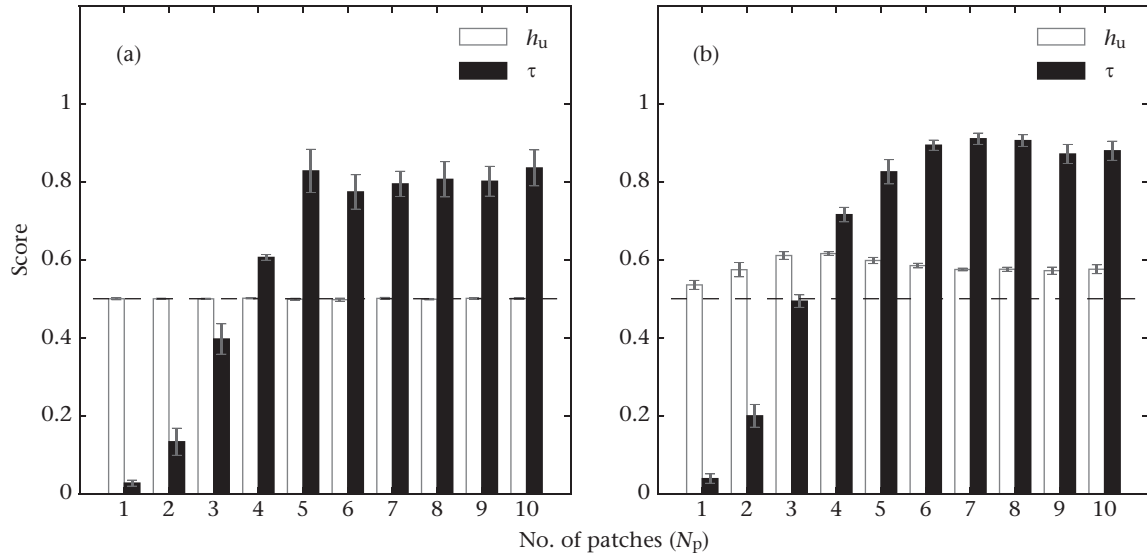


Figure A3. Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches, for the absence of a dominance gradient and high dominance gradient. Parameter values: $c = 1$, $b = 3$, $\gamma = 0.5$, $\lambda = 0.25$, $\delta = 0.1$, $\beta = 1$, $\varepsilon = 1$, $q_0 = 6$, $m = 0.5$, $N = 10$, with patch quality given by equation (3). (a) $d = 0$, (b) $d = 10$. The dashed line shows $h_u = 0.5$.

Effect of benefit of grooming

Fig. A4 illustrates the effect of varying the benefit of grooming (b) on grooming up the hierarchy (h_u). Both decreasing b (Fig. A4a) and increasing b (Fig. A4b), compared to the baseline case in the main text (Fig. 5b), reduces grooming up the hierarchy, thus showing that the hierarchy score is nonmonotonic in b .

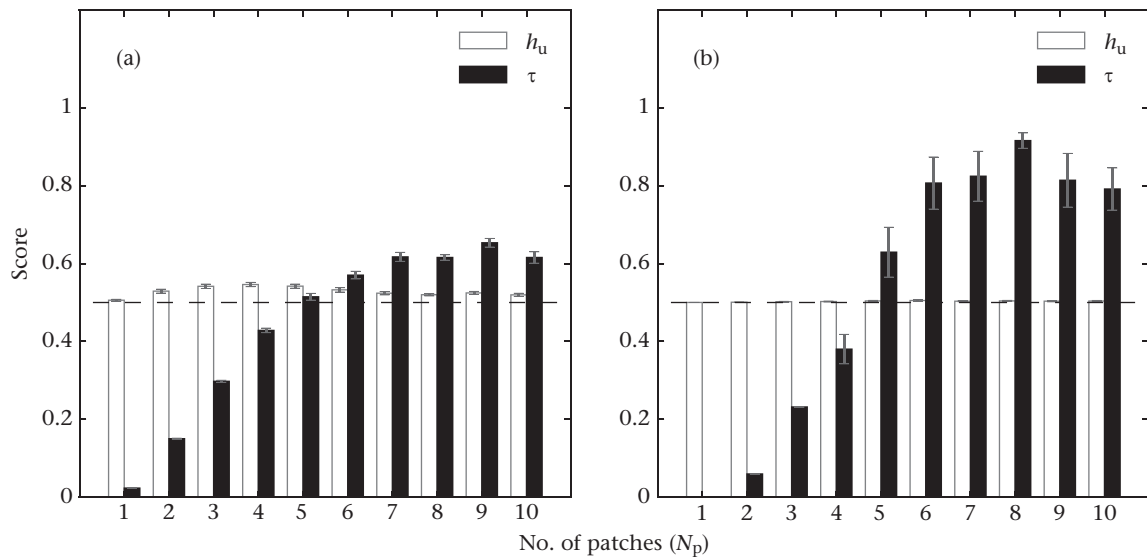


Figure A4. Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches, for low and high benefit of grooming. Parameter values: $c = 1$, $\gamma = 0.5$, $\lambda = 0.25$, $\delta = 0.1$, $\beta = 1$, $\varepsilon = 1$, $d = 2$, $q_0 = 6$, $m = 0.5$, $N = 10$, with patch quality given by equation (3). (a) $b = 1.5$, (b) $b = 4.5$. The dashed line shows $h_u = 0.5$.

Sensitivity to Parameters

Testing patch quality assumptions

Using different assumptions regarding patch quality does not have a qualitative effect on the outcome. We tested both the case where the patch quality is either high or low and the case where all patches are of equal quality (see section ‘Stage 4: Feeding’ from the main text for additional information on our assumptions regarding feeding). Despite alternative assumptions regarding quality, we continue to find that the hierarchy score (h_u) is nonmonotonic in the number of patches (N_p), where the score is maximized for an intermediate number of patches and decreases for more or fewer patches (Fig. A5). Additionally, for $N_p = 10$ when all patches are of equal quality we find that grooming is exactly reciprocal. Here, all individuals will maximize payoff when feeding alone on a patch, and thus there is no incentive to groom for tolerance.

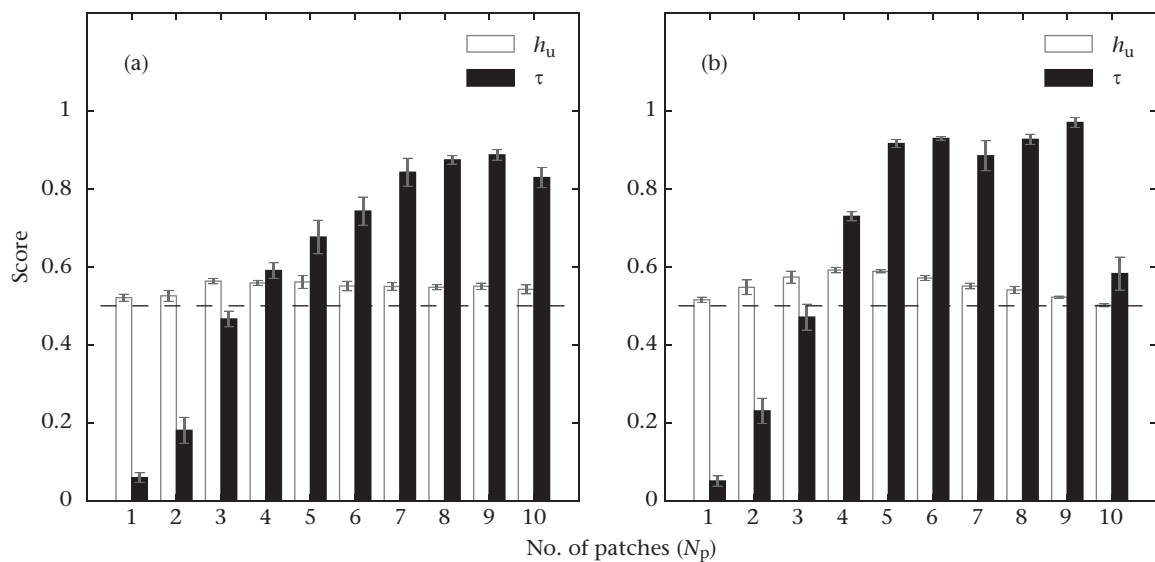


Figure A5. Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches, for different assumptions regarding patch quality. Parameter values: $c = 1$, $b = 3$, $\gamma = 0.5$, $\lambda = 0.25$, $\delta = 0.1$, $\beta = 1$, $e = 1$, $d = 2$, $q_0 = 4$, $m = 0.5$, $N = 10$. (a) High/low patch quality scenario, (b) equal patch quality scenario. The dashed line shows $h_u = 0.5$.

Effect of group size

To determine the robustness of our results we varied the group size (N). We find we can qualitatively reproduce similar patterns for the grooming up the hierarchy score (h_u) for group sizes ranging from five to 100 (Fig. A6). For group sizes $N = 5$, 50 and 100 (and $N = 10$ in the main text), the model produces either grooming up the hierarchy that is maximized for a few patches, a nonmonotonic relationship with number of patches or reciprocal grooming, depending on the level of aggressiveness. Leaving all else equal, however, a larger group size will result in more reciprocal grooming as there will be a smaller proportion of the group feeding on the available patches (Fig. A6).

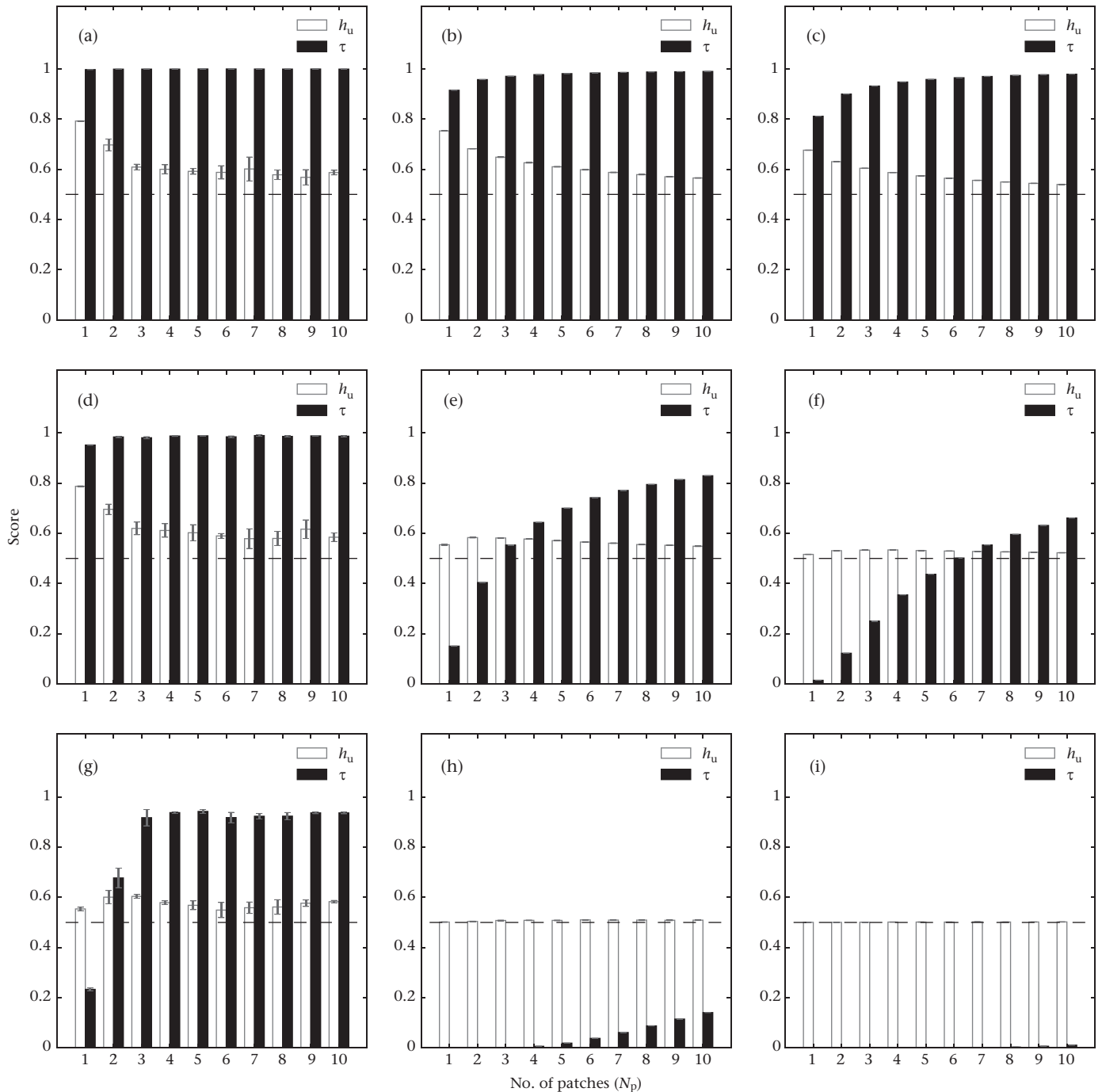


Figure A6. Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches, for different group sizes and levels of aggressiveness. Parameter values: $c = 1, b = 3, \gamma = 0.5, \lambda = 0.25, \delta = 0.1, \epsilon = 1, d = 2, q_0 = 6, m = 0.5$, with patch quality given by equation (3). (a) $N = 5, \beta = -6$, (b) $N = 50, \beta = -6$, (c) $N = 100, \beta = -6$, (d) $N = 5, \beta = -3$, (e) $N = 50, \beta = -3$, (f) $N = 100, \beta = -3$, (g) $N = 5, \beta = 0$, (h) $N = 50, \beta = 0$, (i) $N = 100, \beta = 0$.

Effect of food competition due to m

Whereas the number of patches N_p determines the average number of other individuals a focal individual has to compete with on a food patch, the interference parameter m regulates the level of interference. That is, the larger m is the more food is wasted when multiple individuals are feeding on the same patch. We find that

increasing m results in less grooming up the hierarchy (h_u decreases, Fig. A7a, c). Reducing m has the same effect as reducing q_0 , since it reduces the payoff when individuals co-feed on a patch. If individuals gain less payoff when feeding on a patch, strong feeding associations (high a_{ij}) between individuals are less likely to occur, since individuals will vary their patch choice more. This results in

low-ranking individuals not grooming for tolerance. We find that increasing the average patch quality through q_0 counters the effect of increasing m (Fig. A7b, d).

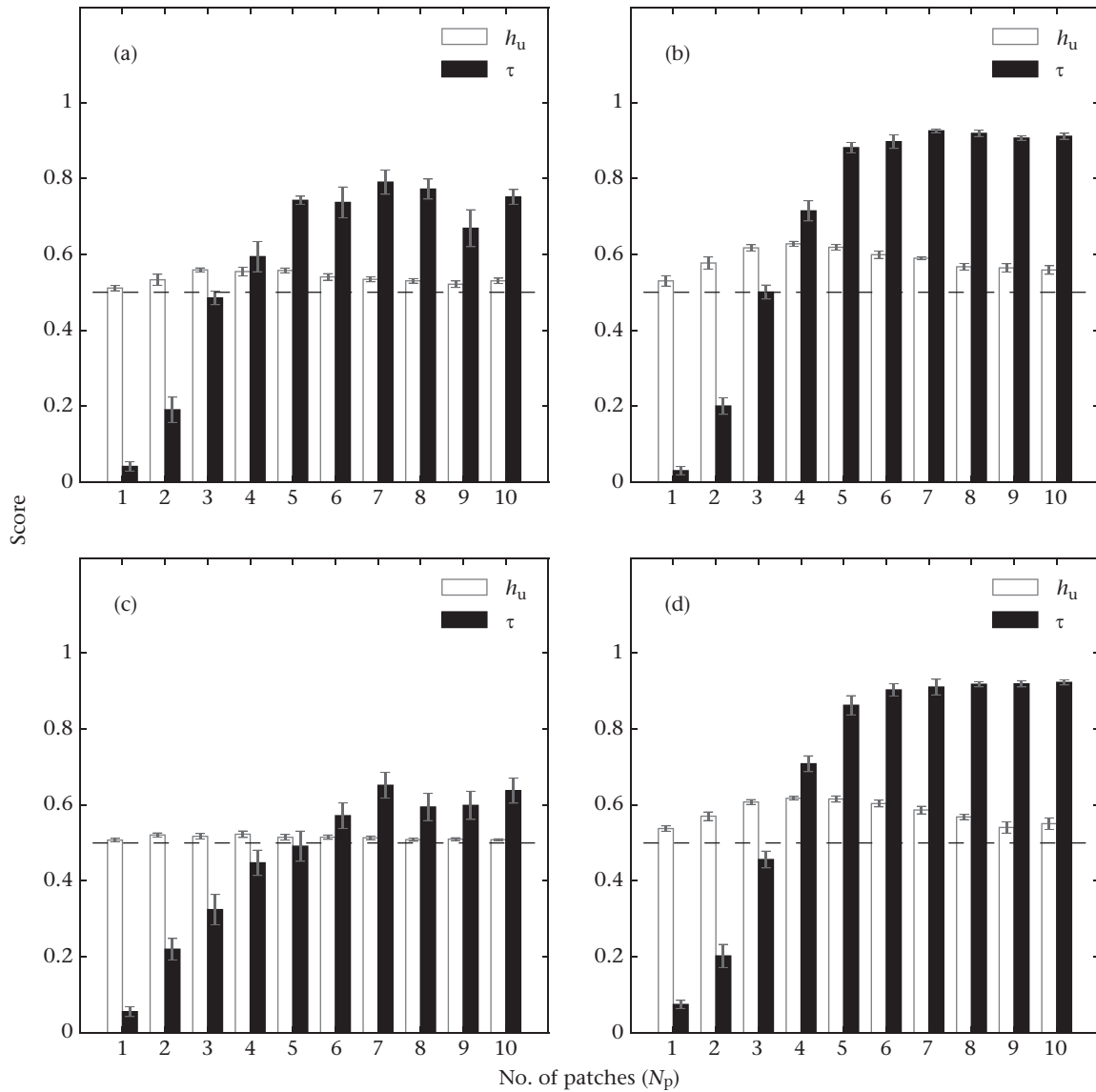


Figure A7. Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches, for different levels of interference and patch quality. Parameter values: $c = 1, b = 3, \gamma = 0.5, \lambda = 0.25, \delta = 0.1, \beta = 1, \epsilon = 1, d = 2, N = 10$, with patch quality given by equation (3). (a) $m = 1, q_0 = 6$, (b) $m = 1, q_0 = 10$, (c) $m = 1.5, q_0 = 6$, (d) $m = 1.5, q_0 = 14$. Whereas increasing m makes grooming more reciprocal ($h_u \rightarrow 0.5$), increasing q_0 cancels the effect. The dashed line shows $h_u = 0.5$.

Cost of grooming

With $c = 0$ grooming another individual will no longer reduce the motivation to groom that individual, and therefore grooming bonds are more easily formed (equation (19)). We thus find reduced levels of grooming up the hierarchy in the absence of a cost of grooming (Fig. A8), although qualitatively the pattern matches our main result (Fig. 5b).

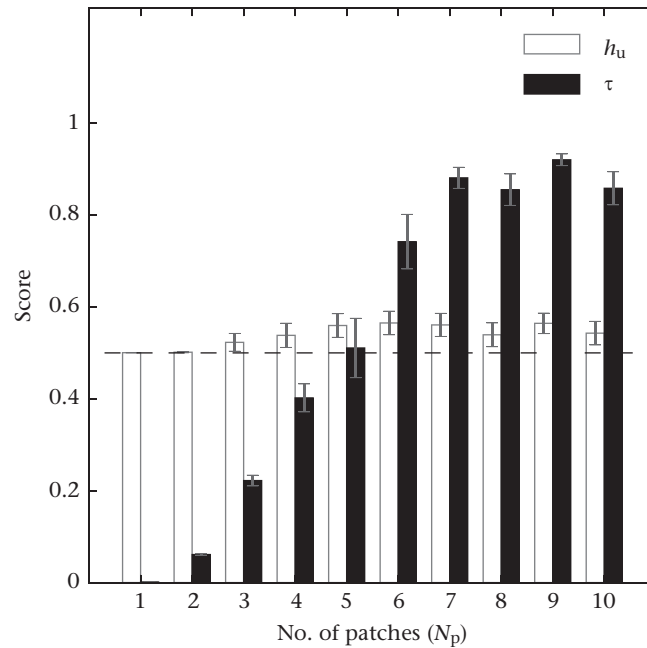


Figure A8. Mean and standard error (over 10 replicates) of the hierarchy score (h_u) and tolerance (τ) as a function of the number of patches, in the absence of a cost of grooming. Parameter values: $c = 0$, $b = 3$, $\gamma = 0.5$, $\lambda = 0.25$, $\delta = 0.1$, $\beta = 1$, $\varepsilon = 1$, $d = 2$, $q_0 = 6$, $m = 0.5$, $N = 10$, with patch quality given by equation (3). The dashed line shows $h_u = 0.5$.

Conditions Favouring Grooming One Rank Up the Hierarchy

To identify conditions where individuals are likely to groom one rank up the hierarchy, we performed several additional simulations where we (rather artificially) created conditions where we would expect such grooming to occur. We used a different equation to compute the payoff per individual on a patch here, where individual payoff drops sharply if more than two individuals feed from the same patch (Fig. A9a, b). Such payoffs will ‘force’ the group to feed in pairs on the available patches. Fig. A9c shows that the proportion of grooming directed to the alpha is still higher than one rank up the hierarchy. On closer inspection we find that the second highest-ranked individual avoids feeding with the alpha here since

its payoff is maximized by feeding on patch $k = 2$ and receiving extra grooming. If it fed on the same patch as the alpha it would have to invest in grooming the alpha and thus, despite the higher payoff from feeding, would get a lower overall payoff. Increasing the payoff difference between patches does increase the proportion of grooming that is directed one rank up the hierarchy (Fig. A9d where payoffs are as in Fig. A9b). However, the proportion of grooming directed to the alpha still slightly outweighs the proportion of grooming one rank up the hierarchy. As such, our model cannot explain grooming of closely ranked individuals very well. Other factors not captured by our model may lead to such grooming occurring more frequently.

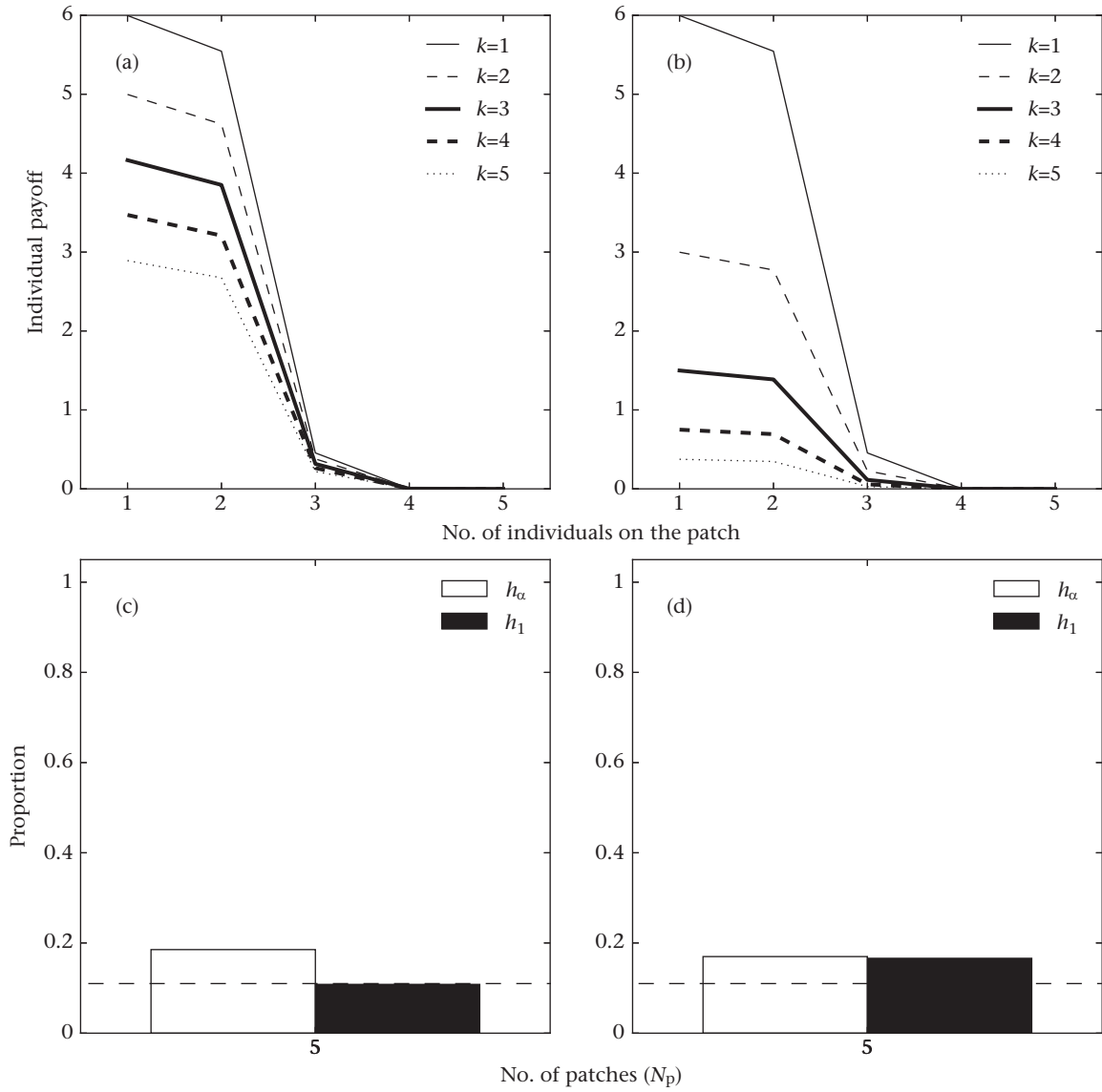


Figure A9. (a, b) Payoff per individual as a function of the number of individuals on a patch, for all five patches. (c, d) Proportion of grooming alpha and proportion grooming one rank up the hierarchy (mean over 10 replicates) with the patch payoffs as in (a) and (b), respectively. Parameter values: $c = 1$, $b = 3$, $\gamma = 0.5$, $\lambda = 0.25$, $\delta = 0.1$, $\beta = 1$, $\epsilon = 1$, $d = 2$, $m = 0.5$, $N = 10$, $N_p = 5$.