22

Cooperation in communication networks: indirect reciprocity in interactions between cleaner fish and client reef fish

REDOUAN BSHARY[1] & ARUN D'SOUZA[2]

[1]*University of Cambridge, Cambridge, UK*
[2]*University of Würzburg, Würzburg, Germany*

Introduction

The aim of this chapter is twofold. First, to outline how recent developments in cooperation theory are so similar to the communication network concept (McGregor, 1993) that a unified terminology would be useful to facilitate exchange of ideas. Second, we argue that the communication network concept provides an evolutionary framework to predict the widespread occurrence of phenomena that until now have been discussed in the context of highly developed cognitive abilities. This creates a problem: as it stands, there appear to be no words in cooperation theory that were not developed in the human context and hence do not include a cognitive component. We have to use definitions that only constitute the functional aspects of phenomena (like tactical deception and indirect reciprocity) and ignore the mechanistic aspects (i.e. theory of mind, intentionality) that are commonly part of the definitions. We kindly ask readers always to keep in mind that our definitions never imply any specific cognitive abilities. We illustrate our ideas with data on the mutualism between the cleaner wrasse *Labroides dimidiatus* and its 'client' reef fish, which visit cleaners to have ectoparasites removed from their surface, gills and mouth (reviewed by Losey *et al.*, 1999; Côté, 2000).

Cooperation provides a challenge to evolutionary theory because it often involves apparently altruistic behaviour. Hamilton (1964) provided a framework to understand why altruism between kin can be evolutionary stable; specifically, an altruist gains indirect fitness benefits from its action. However, there are plenty of

examples where recipients of altruistic acts are unrelated to the helping individual (Dugatkin, 1997). Trivers (1971) pointed out that such altruism might be evolutionarily stable if the recipient later reciprocates, an idea that was formalized by Axelrod & Hamilton (1981). They used an iterated version of the prisoner's dilemma game, a two-player game in which opponents can either cooperate or defect. Defection yields a higher payoff than cooperation independently of the partner's action, but if both players cooperate they receive a higher payoff than if both defect, hence the dilemma. In a computer tournament in which several strategies competed with each other, a simple strategy called 'tit-for-tat' emerged as a cooperative solution to the game. Tit-for-tat players start by being cooperative; then in the next round they play what their opponent did in the previous round. Thus, tit-for-tat players can reap the benefits of mutual cooperation while avoiding exploitation by an uncooperative opponent (except for the first occasion on which an opponent defected). Several new strategies that are similar but apparently superior to tit for tat have been tested since the first computer tournament (reviewed by Dugatkin, 1997).

Despite the intuitive appeal of reciprocal altruism and behavioural strategies similar to tit for tat to ensure cooperative behaviour, few empirical examples have been reported and some that have are contentious (Dugatkin, 1997). In addition, Alexander (1987) pointed out that many examples of human altruistic behaviour do not fit an iterated prisoner's dilemma game: humans often help individuals who are highly unlikely to ever reciprocate. He proposed that humans might help others simply to increase their own image within the society. Nowak & Sigmund (1998) explored this idea by developing a game theory model in which direct reciprocity on altruistic acts was excluded. Instead, an altruist gained an increase in his image score. If an individual's image score was linked to the probability that others were willing to provide help when needed, cooperation readily emerged and was evolutionarily stable (Nowak & Sigmund, 1998). Individuals helped in order to be helped themselves during future interactions with current bystanders (for further theoretical developments see Lotem *et al.* (1999, 2003) and Leimar & Hammerstein (2001). An experiment with first-year students confirmed a crucial prediction of the models: students that helped more than average received more help and, therefore, a final payoff that was above the average (Wedekind & Milinski, 2000). This new approach to the evolution of cooperation is a specific instance of a communication network: the interactions between individuals do not happen in a social vacuum but in the presence of other individuals who eavesdrop and thereby extract relevant information for own future interactions with the actors (McGregor, 1993). As a consequence of eavesdropping, it pays individuals to alter their behaviour, either as a general unconditional response (Johnstone, 2001) or specifically in situations where bystanders are present (audience effects; see Ch. 4).

Nowak & Sigmund (1998) proposed that altruism based on indirect reciprocity is a hallmark of human evolution. Although they do not specify why this should be so, their proposal implies that non-human animals either do not live in social environments that would favour the evolution of altruism through indirect reciprocity or lack some of the cognitive abilities required (but see Zahavi, 1995; Roberts, 1998). In contrast, we suggest that the communication network concept has the advantage of coming from a purely functional perspective rather than trying to explain apparently maladapted human behaviour. Early discussions of communication networks (McGregor, 1993) argued that behaviours such as eavesdropping and audience effects should evolve in the context of a network, without detailed consideration of the underlying cognitive mechanisms. By ignoring mechanisms, we use the communication network concept to predict eavesdropping and audience effects in potentially cooperative contexts without worrying about cognitive constraints. Experimental evidence for the existence of eavesdropping has been provided for a wide variety of taxa (Ch. 2). Whereas human subjects can be asked about their behaviour, eavesdropping in other animals has to be inferred from the eavesdroppers' subsequent behaviour towards individuals observed interacting. Differences in individuals' roles must elicit differences in subsequent eavesdropper behaviour towards them. It was thus an implicit assumption of communication-network studies that eavesdroppers attribute some sort of image score to observed individuals and that this score governs their own future behaviour towards those individuals. Scoring an individual's tendency to help is just one type of image score. Fighting ability, aggressiveness and mating success with regard to female choice are the image scores typically studied in communication networks (e.g. Ch. 5).

That eavesdroppers adjust their own behaviour to what they have witnessed has important implications for the behaviour of individuals that is observed. While a classical approach would suggest that individuals maximize payoffs in each single interaction (with the exception of reciprocal altruism and punishment, where benefits are delayed), any occurrence of eavesdropping implies that selection favours individuals that optimize current actions by integrating both immediate payoffs and future consequences of their behaviour. Within the framework of cooperation theory, it may pay individuals to be altruistic if this increases the probability of meeting more cooperative eavesdroppers in the future; in contest theory, it may pay to be more aggressive in the presence of potential challengers if winning a fight results in fewer attacks from eavesdroppers. Individuals can respond to eavesdropping in two ways. First, they can alter their behaviour in any interaction in relation to the average probability that eavesdroppers are present. In this case, all individuals behave in the same way (with respect to eavesdroppers) in all interactions. Second, individuals can pay attention to specific cues that eavesdroppers

are present for a particular interaction and alter their behaviour accordingly. In the latter, individuals show a flexible behavioural pattern. Communication network models have until now dealt with the first scenario (Nowak & Sigmund, 1998; Johnstone, 2001; Leimar & Hammerstein, 2001). There is, however, increasing evidence that animals adjust their behaviour in a particular interaction according to the presence or absence of eavesdroppers (Doutrelant *et al.*, 2001; Bshary, 2002; Ch. 4).

Altruism towards unrelated individuals has been linked to positive reciprocity, be it direct (Trivers, 1971) or indirect (Nowak & Sigmund, 1998): individual help because they will receive help in return. However, this is not necessarily the case. An alternative is that an individual helps in order to raise its image and uses its image to exploit recipients or eavesdroppers, which will behave cooperatively because of this high image. If all individuals in a population exploited eavesdroppers, then image scoring would break down. However, as long as image scoring yields an overall benefit, either because most altruistic acts are honest in that an individual's willingness to cooperate is revealed or because the benefits of cooperation exceed the costs of being exploited, altruism may be used both as an honest and as a deceptive signal. Therefore, altruism may sometimes serve as a signal out of context, causing other individuals to react in the signaller's favour and to their own disadvantage. This is the functional definition of tactical deception (Hauser, 1998). In communication-network terms, it may pay individuals to be altruistic if this signal is misinterpreted by eavesdroppers in a way that allows future exploitation of them.

Such a functional approach to tactical deception is in strong contrast to the traditional cognitive approach. Though such behaviour has been described, for example in birds (Munn, 1986), tactical deception is often seen as a hallmark of primate 'Machiavellian intelligence' (Byrne & Whiten, 1988): the notion that most primate species have been strongly selected for the cognitive abilities to cope with their social environment (see references in Byrne & Whiten 1988 and Ch. 25). The ability to use tactical deception has, therefore, been linked to the concept of theory of mind (Premack & Woodruff, 1978): the ability to speculate how another individual might perceive a certain situation. However, Heyes (1998) cautioned that any observations of tactical deception do not imply the existence of particularly high cognitive abilities. Instead, originally animals might have made an error (i.e. produced a signal out of context) but it may have had a favourable outcome for the signaller. As a results the signaller may associate this error with a reward and consequently would be more likely to produce the signal again in this context. The notion that simple associative learning might suffice to produce signals that fit the functional definition of tactical deception offers the possibility of using a much more functional, rather than mechanistic, approach to the topic

(see also Hauser, 1998) – and the appropriate framework for the study of tactical deception is communication networks.

Interactions between cleaner fish and clients

In the remainder of this chapter, we present data on interactions between the cleaner wrasse and client reef fish to illustrate the arguments outlined above. Data were collected in the Red Sea, at Ras Mohammed National Park, Egypt. Methods of data collection are described in detail elsewhere (Bshary, 2001, 2002), so here we will keep this kind of information to a minimum. All data are field observations; therefore, experimental proof is still lacking. However, these data illustrate that it is worthwhile searching for potential examples of positive indirect reciprocity and tactical deception with a functional perspective rather than worrying about cognitive constraints.

Clients regularly visit the cleaners at their small territories called 'cleaning stations' (cleaning symbiosis reviewed by Losey *et al.*, 1999; Côté, 2000). As individual cleaner wrasse may have more than 2000 interactions per day (Grutter, 1995), interactions often take place in the presence of other potential visitors. Such bystanders can eavesdrop and evaluate the cleaner's service quality. While the cleaner fish eat parasites, in particular gnathiid isopods (Grutter, 1996), they also feed on client mucus and scales (Randall, 1958; Grutter, 1997). Feeding on healthy client tissue is correlated with the occurrence of client jolts, an observable short shake of the client's body, in response to mouth contact by the cleaner fish (Bshary & Grutter, 2002a). The frequency of client jolts correlates negatively with parasite load; therefore, client jolts are not a byproduct of parasite removal. Rather, jolts are an easily observable correlate of cleaner fish cheating (Bshary, & Grutter, 2002a). Note that only non-predatory clients (i.e. species that could not eat cleaner fish) jolt on a regular basis, while jolts of predatory clients are infrequent (Bshary, 2001). Therefore, we will only present data on non-predatory clients. In response to a jolt, clients often dart off or chase the cleaner, depending on their strategic options. Client species with large home ranges that cover several cleaning stations ('choosy clients') usually make use of their choice options and swim off and visit a different cleaning station for their next inspection (Bshary & Schäffer, 2002), as predicted by biological market theory (Noë *et al.*, 1991; reviewed by Noë, 2001). In contrast, client species with small territories or home ranges, and hence with access to only one cleaning station ('resident clients'), tend to punish cleaners by chasing them (Bshary & Grutter, 2002a). Both darting off and punishment could readily provide bystanders with the information that a cleaning service was bad. In contrast, if an observed interaction ends without apparent conflict, then the service had probably been good. Therefore, clients could easily attribute an image

score to a cleaner fish, and cleaners could adjust their behaviour to the presence of eavesdropping potential clients.

Why should clients attribute an image score to cleaners?

Attributing an image score to an individual and basing behavioural decisions during interactions with that individual on that score only makes sense if the score has some predictive power about how that individual will behave. In potentially cooperative interactions, a positive image score is attributed to an individual only if cooperation on one occasion is usually followed by cooperation on the next occasion. In the context of cleaning symbiosis, this implies that there must be consistent variation in cheating rates either between individual cleaners or within individual cleaners. Indeed, Bshary (2002) found that a minority of cleaners cheated more frequently than the rest. These 'biting cleaners', compared with normal cleaners, specifically targeted larger non-predatory clients, both residents (median client jolt rate was 12/100 seconds in interactions with biting cleaners compared with 2/100 seconds in interactions with normal cleaners) and choosy clients (18/100 seconds compared with 3/100 seconds), while there was no evidence for increased cheating of predatory clients (0/100 seconds compared with 0/100 seconds) or small resident clients (6/100 seconds compared with 6/100 seconds) (Bshary, 2002). These data suggest that it would pay larger non-predatory clients to avoid interactions with such biting cleaners. One way they could do this is to extract information from ongoing interactions and attribute an image to a cleaner.

There is another reason why constant image scoring of cleaner behaviour is advantageous for clients. Data from one biting cleaner fish revealed that cheating rates changed considerably over a period of six weeks. Some of the 99% confidence intervals around observed daily jolt rates of choosy clients did not overlap, suggesting that the variation is significant (Fig. 22.1). This individual was a female, as were all other biting cleaners that have been observed ($n = 7$). Cheating of non-predatory choosy clients peaked at the two periods of full moon that occurred during the observation period, and full moon coincided with repeated spawning with her male partner. After the second spawning period, the male disappeared and cheating rates fell to very low values. The cleaner wrasse is a protogynous hermaphrodite; that is, individuals start their reproductive career as females and eventually switch sex to become males (Robertson, 1972). Males have a larger reproductive output because they often have a harem. Therefore, females face a trade-off between investing in current reproductive effort through the production of eggs and investing in growth to become a male. If the energy requirements for egg production are maximal close to spawning, females needed extra energy in order to avoid compromising growth too much. We suggest that the females' switch to a
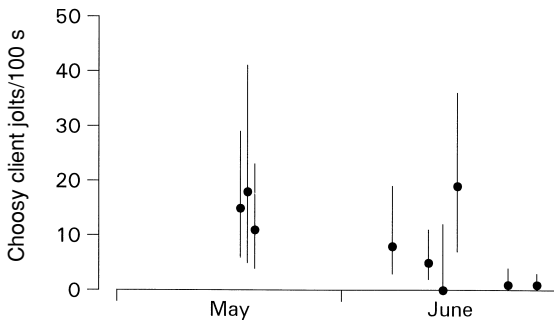
Fig. 22.1. Jolt rates (with 99% confidence intervals) of choosy clients (species with large home ranges that cover several cleaning stations) when interacting with one particular female cleaner on nine different days, based on one hour of observations on each day.

temporarily deceptive strategy yields short-term energetic advantages. In aquaria, clients jolt more frequently when interacting with hungry cleaners that when interacting with satiated cleaners (A. S. Grutter, unpublished data). The benefits of cheating, therefore, seem to be condition dependent, and the client control mechanisms like punishment (Bshary & Grutter, 2002a) and partner switching (Bshary & Schäffer, 2002) only work most of the time.

It even appears that the same individual can switch back and forth between a cooperative and a biting strategy within seconds. Another biting female, observed over a six-week period, cheated clients frequently during the spawning period but client jolt levels remained high after that. Her male partner tolerated her presence at his cleaning station only during spawning but not thereafter and chased her off repeatedly. The female spent about equal amounts of time at her own cleaning station on the other side of the reef patch and on excursions to the male's cleaning station. When the female was at the male's cleaning station, her resident and choosy clients jolted significantly more frequently than when interacting with the female at her own cleaning station (residents ($n = 8$): $t = 1.5$; $p = 0.021$; choosy clients ($n = 7$): $t = 0$; $p = 0.016$; Wilcoxon matched-pair signed-ranks tests; Fig. 22.2). The variation both within and between cleaners apparent from these examples means that client image scoring is a profitable strategy to avoid (temporarily) cheating cleaners.

Do clients attribute image scores to cleaners?

As shown elsewhere (Bshary, 2002), clients use information about the outcome of ongoing interactions when visiting a cleaning station. To appreciate fully what is happening, it is important to note that clients usually do not 'hang out' at cleaning stations but visit them only when they seek an inspection by a cleaner.
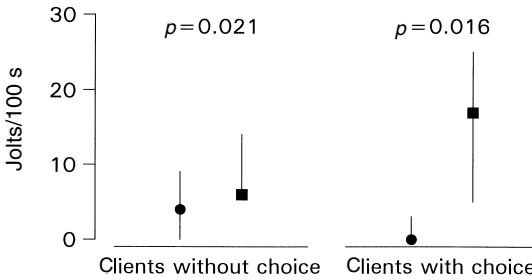
Fig. 22.2. Jolt rates (medians and interquartile ranges) of eight resident species (clients without choice; $n = 8$) and seven client species with choice (defined in text; $n = 7$) during interactions with one particular female cleaner depending on the location of the interaction (at her own station (•) or at the male's station (■); see text for further details).

Therefore, clients can gather information on how cleaners treat other clients only when they visit the station themselves and only if a cleaner is busy inspecting another client while they approach. So visiting clients can base their decision to invite inspection on current information only if they can observe another client being inspected. The newly arrived individual can attribute a positive image score to the cleaner if the current interaction ends without apparent conflict and a negative image score if the current interaction ends with the client darting off or chasing the cleaner. If another client is not present when the client arrives, no current information is available and the image score might be neutral. This is what the data suggest. If an ongoing interaction ended without apparent conflict, clients that had arrived during the interaction invited inspection by the cleaner in almost 100% of observed interactions. In contrast, if the interaction ended with an apparent conflict, clients hardly ever invited inspection (Bshary, 2002). When no information about a cleaner's previous interaction was available, clients invited inspection with intermediate probability and the actual outcome of the previous interaction (that was unobserved by the client) had no significant effect. When clients do not invite inspection, they often exhibit an ambiguous response; they let the cleaner approach and inspect but do not stop coordinated swimming movements before the interaction starts (they may stop afterwards). They may also flee from the approaching cleaner. Fleeing most often results in no inspection and frequently happens when clients are approached immediately after an interaction has ended with a conflict. In contrast, fleeing hardly ever occurs after a positive interaction had just ended or if the previous interaction had ended a while ago (Fig. 22.3). These observations are consistent with the statement above that clients only visit cleaning stations to seek an inspection. The decision to invite inspection is only altered if they observe a negative interaction.
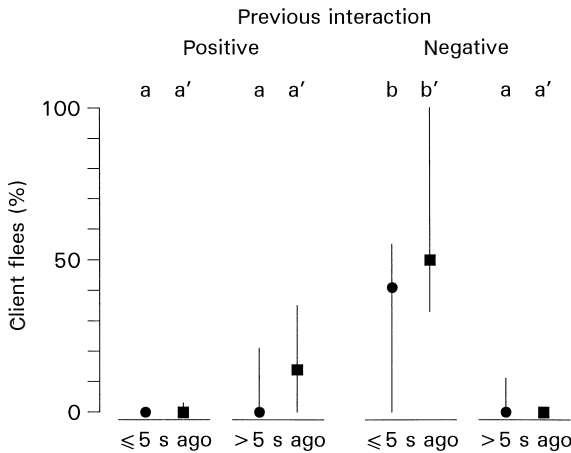
Fig. 22.3. The frequency of fleeing (accelerating away from approaching cleaners) by resident client species (●) and choosy client species (defined in text, ■) arriving at a cleaning station in four different situations: the previous interaction ended either ≤ 5 s or ≤ 5 s ago and had positive (without conflict) or negative (with conflict) outcomes. Values are median and interquartile ranges. The letters b and b′ above the values for fleeing in the situation where the previous interaction had ended negatively ≤ 5 s ago indicate a significant difference to the other three situations, which are not statistically significant between each other, as indicated by using the same letters a and a′.

Response of 'normal' cleaner fish to image-scoring clients

If clients attribute image scores to cleaners, one would expect that cleaners adjust their behaviour and cheat current clients less frequently if bystanders are present than when no bystanders are present. Such audience effects should be particularly common if bystanders have access to several cleaning stations, as these species (see above) might not only delay their interaction with the cleaner but also swim to another cleaning station. Resident bystanders can only delay their interaction or avoid interactions altogether and remain uncleaned. To look for such effects, we assumed that all individuals within 50 cm of cleaner–client interactions at the beginning and at the end of each interaction were able to collect information about the ongoing interaction and all individuals ≥ 10 cm total length were potential next clients. We quantified the number of all such individuals and their species identity for 12 cleaners. For each client species and cleaner station, we calculated correlations between the frequency of jolts and the number of bystanders. We distinguished four classes of bystander; conspecific, heterospecific, resident species and choosy species. We only calculated correlations for observations where only one of these classes of bystander was present. For each client species and bystander category, we compared the number of positive and
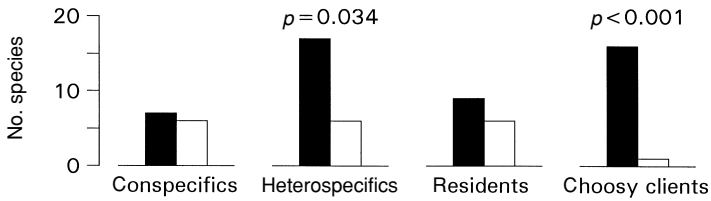
Fig. 22.4. The influence of the presence (within 50 cm at the beginning and the end of an interaction) of four bystander classes on the jolt rate of clients during interactions with cleaners. The histograms show the number of client species for which the correlation between jolt rate and the number of bystanders of a category was either negative (black) or positive (white).

negative correlations and scored a plus if the majority was positive and a minus if the majority was negative. Thus, we had one data point for each client species and bystander category and evaluated any significant impacts of bystander categories on client jolt rates using sign tests. We did not find a significant effect of conspecific bystanders on client jolt rates ($n = 13$; $x = 6$; NS) while the presence of heterospecific bystanders had a significantly negative effect ($n = 23$; $x = 6$; $p = 0.034$; Fig. 22.4). The effect of heterospecific bystanders was mainly owing to choosy bystanders ($n = 17$; $x = 1$; $p < 0.001$) while resident bystanders did not have a significant effect on client jolt rates ($n = 15$; $x = 6$; NS; Fig. 22.4).

While the data presented above are in line with the hypothesis that client image scoring influences cleaner fish behaviour, there is an alternative explanation. It could be that when more clients are present it is easier for cleaners to pick the few obvious parasites from each of them and the reduction in client jolt rate is a result of an optimal foraging decision of cleaners rather than caused by bystander image scoring. In favour of the optimal foraging interpretation, it is known that choosy clients are, on average, larger than resident clients (Bshary 2001), which could explain the stronger effect of their presence on the current clients' jolt rates than the smaller residents. However, optimal foraging cannot explain our observation that choosy bystanders have different effects on cleaner fish behaviour, depending on whether they are the same species as the interacting client or whether they are a different species. We can explore this effect further by considering only interactions in which individuals of one choosy species, the sergeant major *Abudefduf vaigiensis*, were bystanders. We picked 13 cleaning stations for data collection on the basis that these clients were frequent visitors. Sergeant majors may visit as single individuals or as large shoals of 20–50 individuals. For 11 out of 12 client species, we found more negative than positive correlations between the number of sergeant majors present and client jolt rates (sign test: $n = 12$; $x = 1$; $p < 0.01$). This result is the opposite of the effects of sergeant major bystanders on the jolt
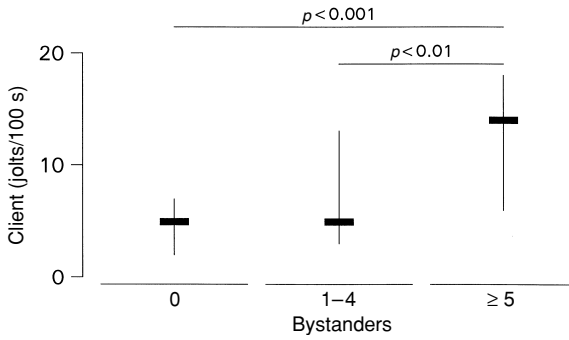
Fig. 22.5. The influence of the number of sergeant major *Abudefduf vaigiensis* bystanders present (within 50 cm at the beginning and the end of an interaction) on the jolt rate of the sergeant major client.

rates of sergeant major being cleaned: in comparison to interactions in which no bystanders were present, small numbers of bystanders did not have any detectable effect and the presence of large numbers led to an increase in jolt rates (Friedman test: $n = 12$; $\chi^2 = 9.9$; df $= 2$; $p < 0.01$; Fig. 22.5).

Response of biting cleaners to image scoring clients

Biting cleaners have more interactions that end with a conflict and clients approaching their cleaning station more often avoid them than normal cleaners (Bshary, 2002). As explained above, it seems likely that the latter observation is the result of client image scoring rather than previous direct experience of clients. Do biting cleaners still have some means to improve their image? In this respect, it is important to note that biting cleaners behave very differently from normal cleaners, not only with respect to jolt rates of large clients but also with respect to their behaviour towards small resident clients. Biting cleaners often ride above the small residents' dorsal area and provide tactile stimulation with their pectoral and pelvic fins. While this behaviour is part of every cleaner's repertoire, about 50% of the interactions between biting cleaners and small residents consisted of tactile stimulation only (Bshary, 2002). Providing tactile stimulation is incompatible with foraging; hence interactions that consist of tactile stimulation only are clearly costly to cleaners. Usually, cleaners provide tactile stimulation in response to the behaviour of the client; for example, manipulating clients that are unwilling to interact. The manipulation serves to slow down the clients, allowing the cleaners to forage on the clients' surface (Bshary & Würth, 2001). As tactile stimulation of small residents did not appear to provide the cleaners with any direct benefits from the recipients, Bshary (2002) proposed that it may serve as a signal to attract image-scoring clients, which can then be exploited. In line with this argument, it was found that interactions that consisted of tactile stimulation only were followed
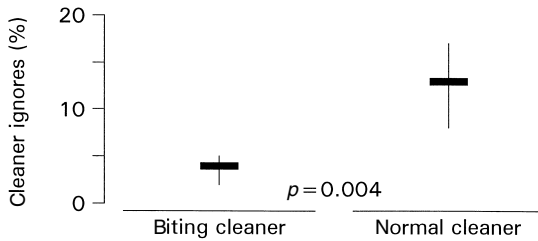
Fig. 22.6. Frequencies with which biting and normal cleaners ignored the invitations for inspection of small resident clients. Values are the median and interquartile ranges for five biting and eleven normal cleaners. (Mann–Whitney U-test: for p value; see text.)

by interactions ending with a conflict immediately after a client jolt more often than expected. It appears that tactile stimulation of small residents is a signal out of context that attracts image-scoring clients to their own disadvantage (they will be cheated) and to the cleaners' advantage, fulfilling the functional definition of tactical deception (Hauser, 1998).

The presence or absence of bystanders was not noted, so it remains unclear whether biting cleaners seek small residents in particular when larger clients are nearby or whether they start such interactions independently of the presence of bystanders. The latter scenario is more plausible, as larger clients, in particular the choosy ones, are not willing to queue for inspection (Bshary & Schäffer, 2002) and would, therefore, swim off despite the cleaner's positive image. So, cleaners appear unable to time interactions with small residents for maximal effects. However, some evidence suggests that tactile stimulation of small residents is part of biting cleaners' strategies to improve their image. Cleaner fish sometimes ignore clients that invite inspection, in particular small resident clients (Bshary & Würth, 2001). These clients do not offer a large food source and do not have the option to visit another cleaner if ignored. When approached by small residents, biting cleaners ignore them significantly less frequently than normal cleaners (Mann-Whitney U-test: $n = 11$; $m = 5$; $U = 2$; $p = 0.004$; Fig. 22.6).

Discussion

We have provided a description of behavioural patterns in interactions between cleaner fish and client reef fish that emphasizes the importance of the communication-network framework in understanding the dynamics of cooperative interactions and the occurrence of tactical deception. Cleaning interactions often occur in the presence of other potential clients of cleaners. These bystanders eavesdrop on ongoing interactions, and the information that they collect appears

to be crucial for their decision to invite inspection or to avoid the cleaner. In response, it appears that normal cleaners reduce cheating frequencies in the presence of eavesdroppers, in particular if these eavesdroppers have access to several cleaning stations. Data of this kind are still missing for biting cleaners. Biting cleaners frequently engage in costly (or at least non-profit) interactions with small residents that appear to serve to attract larger image-scoring clients, which can then be exploited. The results have important implications for theoretical approaches to indirect reciprocity. Existing models predict that image scoring drives altruistic behaviour towards fixation (Nowak & Sigmund, 1998; Lotem *et al.*, 1999; Leimar & Hammerstein, 2001). Cheating individuals can only reinvade an image-scoring population after genetic drift has led to an increase in non-discriminatory altruists. This scenario does not fit the cleaner fish mutualism very well. Image scoring of clients mediates cooperative behaviour of cleaners, but this cooperative behaviour may be an honest or a deceptive signal. Cheating individual cleaners use one class of clients for altruistic behaviour to produce a signal that allows them to exploit, through image scoring, another class of clients. Image scoring thus works for the receiver of altruistic behaviour but it does not always work for the eavesdropper.

The major difference between the cleaner fish system and the models might concern the payoff matrix. While it is assumed in the models that payoffs are the same in every interaction, payoffs are variable for cleaners. First, an advantage of cheating clients with access to several cleaning stations rather than resident clients is that the former just swim off after being cheated, while the latter chase the cleaner fish around (Bshary & Grutter, 2002a), so the cleaner loses some of the energy it has just gained. Second, cleaners can probably gain very little from interactions with small clients anyway, no matter whether they cooperate or cheat. This contrasts with interactions with large clients, which have more parasites but also more mucus and a larger surface for the cleaner to scrape along with its lower jaw. This gives the opportunity for cleaners to behave altruistically when payoffs are low and to be exploitative when payoffs are high, as long as the altruism increases the frequency of high payoff interactions. Image scoring would not persist if it did not yield a benefit to its performer, but it also provides an opportunity for individuals to perform altruistic acts in order to gain access to and exploit image-scoring individuals. Therefore, image scoring in communication networks may explain both the evolution of altruistic behaviour and the occurrence of tactical deception. The commonness of dishonest signals that nevertheless still fool observers has yet to be evaluated. While verbal arguments predicted low frequencies (Dawkins & Krebs, 1978), game theoretic models indicate that this is not necessarily the case (Johnstone & Grafen, 1993; Szamado, 2000). In particular, if the benefits of finding a cooperative partner largely outweigh the cost of

interacting with a cheating partner, tactical deception may occur at quite high frequencies.

Future work with the cleaner system

Several important points of the cleaner fish symbiosis still have to be clarified. First, we need experimental evidence for both client image scoring and cleaner fish audience effects. There is increasing evidence that a reduction in client jolt rates in the presence of bystanders reflects a more cooperative behaviour by cleaners (Bshary & Grutter, 2002b) and that such behaviour is indeed more altruistic. A. S. Grutter & R. Bshary (unpublished data) offered cleaners the choice between equal amounts of mucus, gnathiid isopods and monogeneans attached to plexiglas plates and found that cleaners ate mucus more often than parasites, in particular gnathiids. Assuming that the results reflect the items' quality as a food source, cleaners profit even more from feeding on mucus when interacting with real clients as mucus is abundantly spread over the clients' surface whereas parasites have to be searched for. In conclusion, while the experiment did not quantify energy intake, it makes it very plausible that feeding on mucus yields a higher energy gain than feeding on parasites. Another point that needs to be addressed is the biting cleaners' behaviour with respect to small and large clients. Is it really true that interactions with small clients generally offer low payoffs compared with interactions with larger clients, *and* that the margin between the benefits from cooperation and cheating increase with client size? Does image scoring of clients indeed inhibit the cleaners' tendency to cheat in low-payoff interactions but not in high-payoff interactions? Currently, no data are available to evaluate these questions. Finally, one might expect that clients should respond to the biting cleaners' behaviour by fine-tuning their image scoring, paying less attention to the outcome of interactions between cleaners and small residents. Pooling of existing data indicate that this is indeed the case. Invitation for inspection (i.e. spreading the pectoral fins and stopping coordinated swimming movements) occurred more frequently if cleaners interacted with choosy clients than with small residents. This preliminary result has to be tested with a larger data set that allows statistical analysis based on behaviour of individual client species.

Cognitive aspects

While indirect reciprocity and tactical deception were considered to be a hallmark of human evolution (Nowak & Sigmund, 1998) and primate Machiavellian intelligence (Byrne & Whiten, 1988), the data on cleaner–client interactions suggest that, on a purely descriptive level ignoring underlying mechanisms, these phenomena are more widespread. We have argued that they should occur

frequently in social networks. These phenomena should occur if it pays to alter the optimal behaviour in a situation in order to alter one's image, which will, in turn, produce benefits during future interactions with bystanders that exceed the momentary costs. With respect to aggression, a game theoretic model shows that it may even pay individuals to act spitefully towards a partner (in the sense that the spiteful act will not lead to any benefits gained from future behaviour of the recipient) if this spiteful act reduces, for example, the threat of attack from bystanders (Johnstone & Bshary, 2004; Ch. 10).

With respect to cognition, the data generally support the view of Heyes (1998) that we need to establish what kind of information animals use for their decision making to find out what cognitive abilities are involved in a given phenomenon. Indirect reciprocity and tactical deception may be something 'smart' in some species and simple conditioning in others. Cleaners certainly have ideal conditions to develop their behaviour through conditioning. They have more than 2000 interactions per day (Grutter, 1995), making it easy to connect altruistic behaviour with reward (i.e. the invitation from bystanders to inspect) and cheating with punishment (i.e. evasive actions of bystanders when approached by the cleaner). In the absence of decisive experiments, it could even be possible that parts of cleaner fish and client behaviour may be governed by endocrine responses rather than through learning (Ch. 21). A good candidate for an endocrine-mediated behaviour might be the good service that cleaners provide to predatory clients: there might be an innate programme to recognize predators, and the presence of a predator might trigger a stress response that, in turn, may inhibit cheating behaviour. Alternatively, one could also generalize the Machiavellian intelligence hypothesis and predict that a complex social network should have similar effects on cognitive abilities in all species (Byrne & Whiten, 1988). It may turn out that cleaners have relatively high cognitive abilities, as their large interspecific social network is at least in part based on individual recognition (Tebbich *et al.*, 2002) and demands the solving of a variety of problems (Bshary *et al.*, 2002). In this context, it is worth pointing out that both the biting females which were observed over longer time periods showed considerable variation in their behaviour, as so did their male partners. One male often prevented his female from interacting with clients while the other did not (Fisher test: $n = 34$; $p = 0.003$; Table 22.1). The preventive male was also almost significantly more likely to chase his female when their client darted off after a jolt, while the other male often followed the client to provide tactile stimulation (Fisher test: $n = 12$; $p = 0.053$; Table 22.1). This observation of flexibility of both males and females is important as it was the careful description of individual-specific strategies in primates that eventually led to a cognitive, rather than genetic, approach towards behaviour (Strum *et al.*, 1998).

Table 22.1. *The behaviour of two males that were partners of biting females*

| | Male responses to female biting | | | |
	Tactile stimulation of client[a]	Chasing female[a]	No obvious reaction[a]	Preventative chasing of female[b]
First male	5	2	3	1
Second male	0	5	2	16

[a] Reaction to clients darting off following cheating by female.

[b] Keeping female away from clients.

Summary

In summary, we think that the concept of communication networks has major implications for our understanding of the evolution and maintenance of altruistic behaviour, tactical deception and spiteful behaviour. Because of its functional approach, the communication-network framework may help to demystify phenomena that are often considered to demand high cognitive abilities, opening the way to focus on the underlying mechanisms and the complexity of information processing and decision rules in order to illuminate cognitive differences between species (for a parallel discussion, see Ch. 24). Game theory models should help to generate testable predictions of the circumstances in which altruism, tactical deception and spiteful behaviour may yield fitness benefits within communication networks. In particular, it is time to develop cognitive models rather than genetic models, allowing individuals to process information about their social environment before making a behavioural decision (see Stephens & Clements (1998) for a first approach towards cognitive game theory).

Acknowledgements

References

Alexander, R. D. 1987. *The Biology of Moral Systems*. New York: Aldine de Gruyter.

Axelrod, R. & Hamilton, W. D. 1981. On the evolution of co-operation. *Science*, **211**, 1390–1396.

Bshary, R. 2001. The cleaner fish market. In: *Economics in Nature*, ed. R. Noë, J. A. R. A. M. van Hooff & P. Hammerstein. Cambridge, UK: Cambridge University Press, pp. 146–172.

  2002. Biting cleaner fish use altruism to deceive image scoring clients. *Proceedings of the Royal Society of London, Series B*, **269**, 2087–2093.

Bshary, R. & Grutter, A. S. 2002a. Experimental evidence that partner choice is the driving force in the payoff distribution among cooperators or mutualists: the cleaner fish case. *Ecology Letters*, **5**, 130–136.

  2002b. Asymmetric cheating opportunities and partner control in a cleaner fish mutualism. *Animal Behaviour*, **63**, 547–555.

Bshary, R. & Schäffer, D. 2002. Choosy reef fish select cleaner fish that provide high service quality. *Animal Behaviour*, **63**, 557–564.

Byrne, R. W. & Whiten, A. 1988. *Machiavellian Intelligence*. Oxford: Clarendon Press.

Bshary, R. & Würth, M. 2001. Cleaner fish *Labroides dimidiatus* manipulate client reef fish by providing tactile stimulation. *Proceedings of the Royal Society of London, Series B*, **268**, 1495–1501.

Bshary, R., Wickler, W. & Fricke, H. 2002. Fish cognition: a primate's eye view. *Animal Cognition*, **5**, 1–13.

Côté, I. M. 2000. Evolution and ecology of cleaning symbioses in the sea. *Oceanography and Marine Biology Annual Review*, **38**, 311–355.

Dawkins, R. & Krebs, J. R. 1978. Animal signals: information or manipulation? In: *Behavioural Ecology: An Evolutionary Approach*, ed. J. R. Krebs & N. B. Davies. Oxford: Blackwell, pp. 282–309.

Doutrelant, C., McGregor, P. K. & Oliveira, R. F. 2001. The effect of an audience on intrasexual communication in male Siamese fighting fish, *Betta splendens*. *Behavioral Ecology*, **12**, 283–286.

Dugatkin, L. A. 1997. *Cooperation among Animals: A Modern Perspective*. Oxford: Oxford University Press.

Grutter, A. S. 1995. Relationship between cleaning rates and ectoparasite loads in coral reef fishes. *Marine Ecology Progress Series*, **118**, 51–58.

  1996. Parasite removal rates by the cleaner wrasse *Labroides dimidiatus*. *Marine Ecology Progress Series*, **130**, 61–70.

  1997. Spatio-temporal variation and feeding selectivity in the diet of the cleaner fish *Labroides dimidiatus*. *Copeia*, **1997**, 346–355.

Hamilton, W. D. 1964. The genetical evolution of social behaviour. *Journal of Theoretical Biology*, **7**, 1–52.

Hauser, M. D. 1998. Minding the behaviour of deception. In: *Machiavellian Intelligence II*, ed. A. Whiten & D. W. Byrne. Cambridge, UK: Cambridge University Press, pp. 112–143.

Heyes, C. M. 1998. Theory of mind in non human primates. *Behavioral and Brain Sciences*, **21**, 101–148.

Johnstone, R. A. 2001. Eavesdropping and animal conflict. *Proceedings of the National Academy of Sciences, USA*, **98**, 9177–9180.

Johnstone, R. A. & Bshary, R. 2004. The evolution of spite through indirect reciprocity. *Proceedings of the Royal Society of London, Series B*, in press.

Johnstone, R. A. & Grafen, A. 1993. Dishonesty and the handicap principle. *Animal Behaviour*, **46**, 759–764.

Leimar, O. & Hammerstein, P. 2001. Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London, Series B*, **268**:745–753.

Losey, G. C., Grutter, A. S., Rosenquist, G., Mahon, J. L. & Zamzow, J. P. 1999. Cleaning symbiosis: a review. In: *Behaviour and Conservation of Littoral Fishes*, ed. V. C. Almada, R. F. Oliveira & E. J. Goncalves. Lisbon: Instituto Superior de Psicologia Aplicada, pp. 379–395.

Lotem, A., Fishman, M. A. & Stone, L. 1999. Evolution of cooperation between individuals. *Nature*, **400**, 226–227.

2003. From reciprocity to unconditional altruism through signaling benefit. *Proceedings of the Royal Society of London, Series B*, **270**, 199–205.

McGregor, P. K. 1993. Signalling in territorial systems: a context for individual identification, ranging and eavesdropping. *Philosophical Transactions of the Royal Society of London, Series B*, **340**, 237–244.

Munn, C. A. 1986. Birds that 'cry wolf'. *Nature*, **319**, 143–145.

Noë, R. 2001. Biological markets: partner choice as the driving force behind the evolution of cooperation. In: *Economics in Nature*, ed. R. Noë, J. A. R. A. M. van Hooff & P. Hammerstein. Cambridge, UK: Cambridge University Press, pp. 92–118.

Noë, R., van Schaik, C. P. & van Hooff, J. A. R. A. M. 1991. The market effect: an explanation for pay-off asymmetries among collaborating animals. *Ethology*, **87**, 97–118.

Nowak, M. A. & Sigmund, K. 1998. Evolution of indirect reciprocity by image scoring. *Nature*, **393**, 573–577.

Premack, D. & Woodruff, G. 1978. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, **4**, 515–526.

Randall, J. E. 1958. A review of labrid fish genus *Labroides*, with descriptions of two new species and notes on ecology. *Pacific Scientist*, **12**, 327–347.

Roberts, G. 1998. Competitive altruism: from reciprocity to the handicap principle. *Proceedings of the Royal Society of London, Series B*, **265**, 427–431.

Robertson, D. R. 1972. Social control of sex reversal in a coral-reef fish. *Science*, **177**, 1007–1009.

Stephens, D. W. & Clements, K. C. 1998. Game theory and learning. In: *Game theory and Animal Behaviour*, ed. L. A. Dugatkin & H. K. Reeve. Oxford: Oxford University Press, pp. 239–260.

Strum, S. C., Forster, D. & Hutchins, E. 1998. Why machiavellian intelligence may not be machiavellian. In: *Machiavellian Intelligence II*, ed. A. Whiten & D. W. Byrne. Cambridge, UK: Cambridge University Press, pp. 50–85.

Szamado, S. 2000. Cheating as a mixed strategy in a simple model of aggressive communication. *Animal Behaviour*, **59**, 221–230.

Tebbich, S., Bshary, R. & Grutter, A. S. 2002. Cleaner fish *Labroides dimidiatus* recognise familiar clients. *Animal Cognition*, **5**, 139–145.

Trivers, R. L. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology*, **46**, 35–57.

Wedekind, C. & Milinski, M. 2000. Cooperation through image scoring in humans. *Science*, **288**, 850–852.

Zahavi, A. 1995. Altruism as a handicap : the limitations of kin selection and reciprocity. *Journal of Avian Biology*, **26**, 1–3.