Male cleaner wrasses adjust punishment of female partners according to the stakes
Nichola J. Raihani1*, Ana I. Pinto2, Alexandra S. Grutter3, Sharon Wismers4 and Redouan Bshary2

1Institute of Zoology, Zoological Society London, Regent’s Park, London NW1 4RY, UK
2Université de Neuchâtel, Institut de Biologie, Eco-Éthologie, Rue Émilie-Argand 11, 2000 Neuchâtel, Switzerland
3School of Biological Sciences, The University of Queensland, Brisbane, Queensland 4073, Australia
4Institute of Evolutionary Biology and Environmental Studies, University of Zürich, Winterthurestrasse, 8057 Zürich, Switzerland

Punishment is an important deterrent against cheating in cooperative interactions. In humans, the severity of cheating affects the strength of punishment which, in turn, affects the punished individual’s future behaviour. Here, we show such flexible adjustments for the first time in a non-human species, the cleaner wrasse (Labroides dimidiatus), where males are known to punish female partners. We exposed pairs of cleaners to a model client offering two types of food, preferred ‘prawn’ items and less-preferred ‘flake’ items. Analogous to interactions with real clients, eating a preferred prawn item (‘cheating’) led to model client removal. We varied the extent to which female cheating caused pay-off reduction to the male and measured the corresponding severity of male punishment. Males punished females more severely when females cheated during interactions with high value, rather than low value, model clients; and when females were similar in size to the male. This pattern may arise because, in this protogynous hermaphrodite, cheating by similar-sized females may reduce size differences to the extent that females change sex and become reproductive competitors. In response to more severe punishment from males, females behaved more cooperatively. Our results show that punishment can be adjusted to circumstances and that such subtleties can have an important bearing on the outcome of cooperative interactions.

Keywords: cleaner fish; cooperation; Labroides dimidiatus; mutualism; punishment

1. INTRODUCTION
Punishment can help to sustain cooperation where it would otherwise fail [1]. It is thought to be of particular importance in public goods games in humans, where several studies have shown that people will often pay a fee to reduce the income of uncooperative group members [2–4] (but see [5]). Laboratory experiments have demonstrated that players adjust their investment in punishment according to the deviation between free-riders’ contributions and the mean group contribution, imposing greater penalties on the worst offenders [2–4]. Furthermore, punishment effectiveness (the fee to fine ratio) affects cooperation, with contributions to public goods increasing only when punishment requires a low fee from punishers while imposing a high fine on punished individuals [4,6]. However, the intuitive prediction that harsher punishment precipitates greater cooperation has rarely been formally tested. Moreover, while potential cases abound [1], conclusive evidence for punishment in non-human species is rare [7–9] and evidence for adjusted punishment and corresponding behavioural responses of targets absent. Here, we test these issues using a well-known marine cleaning mutualism between the bluestreak cleaner wrasse, Labroides dimidiatus (hereafter ‘cleaner fish’ or ‘cleaners’), and its reef-fish clients [10].

Cleaner fish live on coral reefs throughout the tropical Indo-Pacific [11]. Male cleaners each hold a territory which encompasses that of several female breeding partners [12,13]. Cleaners provide a service to clients by removing ectoparasites. However, there is a conflict since cleaners prefer to cheat by eating client mucus, which is detrimental to the client [14]. Therefore, in order to receive a good cleaning service, clients require cleaners to cooperate by feeding against their preference. Clients achieve this either by avoiding cleaners they observe cheating other clients, avoiding cleaning stations where they have received a poor service in the past, or aggressively punishing cheating cleaners [8,15]. Conflict may also arise within male–female cleaner pairs, since male–female pairs repeatedly inspect client fish jointly [16]. During pairwise inspections of joint clients, cleaners face a problem akin to a Prisoner’s Dilemma as only one of the pair can obtain the benefit associated with cheating (biting the client) while the cost (client departure) is shared between both partners [16]. Despite the benefits of cheating, cleaners apparently find a cooperative solution to this dilemma, largely because females behave more cooperatively during pairwise than during solitary inspections [16]. This increase in female cooperativeness arises because males punish females that cheat and cause joint clients to leave [7]. Males punish females by aggressively chasing them following model client removal. Conversely, females do not punish males [7]. Whether the severity of cheating leads to corresponding levels of
punishment and whether the punished individuals’ future behaviour depends on punishment intensity remain untested.

In addition to losing a client, a likely reason for male punishment of cheating females in *L. dimidiatus* is that cheating allows females to derive foraging benefits while, at the same time, depriving males of further foraging opportunities with that client. As a consequence, cheating females may increase their food intake (and growth rate) relative to that of their male partner, which would precipitate conflict over reproductive control since females may change sex to become a reproductive competitor if they approach or exceed the male partner in size [11,12,17]. Indeed, in Caribbean cleaner gobies (*Elegantus* spp.), where sex change is absent, males do not punish females that cheat during joint client inspections [18]. The potential for female sex change means that male cleaner wrasses benefit from maintaining a size advantage over their female partner and this generates an additional conflict of interest between males and females over female cheating during joint client inspections. This is reminiscent of findings in other species with size-based dominance hierarchies, where the degree of conflict between same-sex neighbours in the hierarchy and the dominant’s ability to maintain its rank, are tightly linked to the size asymmetry between them [19–21]. In this study, we predicted that conflict over female cheating during joint client inspections would be exacerbated in cleaner fish pairs where the female was close in size to the male partner. This is because the effect of the female obtaining greater nutritional benefits from cleaning interactions than the male would be most detrimental to the male partner when more food might allow the female to out-grow him and therefore change sex.

In the laboratory, the decision rules underpinning cleaner service quality can be investigated using model Plexiglas ‘clients’ containing two food types: prawn (preferred by cleaners) and fish flake (less preferred) [16]. Cleaners are trained that eating prawn results in the immediate removal of the plate such that—as under natural conditions [14]—cooperative behaviour when interacting with model clients also requires cleaners to feed against their preference. With this experimental paradigm, a cleaner that eats a prawn item can be said to cheat its partner: it gains the benefit of eating a preferred food item while its partner is deprived of further foraging. We used this approach to show experimentally that male punishment of cheating females varied with the seriousness of the female’s crime, and, that females reacted to more severe punishment with a greater increase in cooperative behaviour.

### 2. MATERIAL AND METHODS

This research was conducted at Lizard Island Research Station, Australia (14° 40′S, 145° 28′E) in August 2010. Cleaner fish were caught with a barrier net in nearby lagoons and housed in pairs in aquaria (minimal size 45 × 30 × 25 cm) for two weeks before experiments began. The total length (TL) of each fish was measured to the nearest millimetre. Cleaners were each provided with PVC shelter tubes (1 × 10 cm). All fish were trained to feed off Plexiglas plates which had small amounts (items) of mashed prawn or fish flakes mixed with prawn (hereafter ‘flake’) placed on them, and to learn that eating prawn led to the removal of the plate, such that all fish ate flake items before eating prawn items before commencing experiments (see [16] for more details). In experiment 1, two types of plate were used: high value and low value plates. Only high value plates were used in experiment 2. High value plates (15 × 10 cm) contained eight flake items placed within eight black circles (diameter 0.5 cm) and two prawn items placed within two black triangles (0.5 × 0.5 × 0.5 cm) on the plate (see [16] for a schematic). Low value plates were smaller (10 × 5 cm) and contained four flake items placed within four black circles and two prawn items placed within two black triangles (same dimensions as high value plates). All data were analysed using R v. 2.8.1 ([www.r-project.org](http://www.r-project.org)) using linear regression or (generalized) linear mixed models (GLMM) where appropriate. All tests were two-tailed and data examined to ensure the assumptions of statistical tests were satisfied. All two-way interactions were included in the models (where appropriate) but were excluded from the final model if they did not significantly improve the fit of the model to the data.

#### (a) Experiment 1

Data were collected on 12 cleaner fish pairs, though data for one pair were not available owing to the very small size asymmetry within the pair (0.3 cm) making it difficult to reliably distinguish the male from the female. Each pair was observed during 56 trials which were conducted over 5 days. A trial consisted of a single plate presentation to a cleaner fish pair. Cleaners were allowed to feed on the plate until one of the pair ate a prawn item—at this point the plate was removed. We recorded which of the pair ate the prawn item and—if the female ate the prawn item—the intensity of male punishment (number of aggressive chases) in the 30 s following plate removal. Previous work has shown that males are more aggressive in response to female cheating than if males themselves cause plate removal by eating prawn [7,16]. Data from trials where females cheated (by eating the prawn item) were used to investigate how relative female size (female TL/male TL) and client value (high or low) affected male punishment intensity. Mean punishment intensity (per male) was calculated for each client type (high value and low value) and then the mean difference in punishment intensity was calculated (mean punishment intensity (high value clients)—mean punishment intensity (low value clients)).

This approach exploited the paired nature of our data and resulted in a sample size of 11 for analysis. The mean difference in punishment intensity was set as the response term in a linear regression and relative female size was included as the explanatory term.

#### (b) Experiment 2

The second experiment was designed to explicitly test whether male punishment of cheating females increased with relative female size, and whether females behaved more cooperatively in response to more severe punishment. To measure the effect of relative female size on punishment and subsequent female foraging behaviour, we created ‘similar-sized’ pairs and ‘dissimilar-sized’ pairs using a total of 11 females and 15 males. Ten of the females and nine of the males were individuals from experiment 1. The remaining fish were individuals that had been caught for use in another experiment. These individuals were chosen so as to achieve both the minimum (for similar-sized pairs) and the

Punishment fits the crime in fish  N. J. Raihani et al. 367

Figure 1. The mean difference in male punishment intensity (chases per second) when females cheated high value, rather than low value, clients, according to the relative size of the female. The positive slope indicates that male punishment intensity increased when similar-sized females cheated high value clients. The line and confidence intervals are based on the predicted values of the linear regression. Points are means generated from raw data (\(n=11\) cleaner pairs).

maximal possible size asymmetries (for dissimilar-sized pairs) within experimental pairs, respectively. Each female was paired with a dissimilar-sized male (mean size difference (\(\pm\) s.e.) \(= 1.4\pm 0.1\) cm) and with a similar-sized male (mean size difference \(= 0.6\pm 0.1\) cm) to create 11 similar-sized male–female pairs and 11 dissimilar-sized pairs. Data from one of the similar-sized pairs could not be used because the female became behaviourally dominant to the male partner, and the fish were then separated from one another to prevent escalating aggression. Females were paired with each male for a day prior to experiments to allow all fish to recover from the disturbance of being moved to another tank. We always paired females with unfamiliar males to rule out the possibility that familiarity with a social partner could influence behaviour.

Cleaner foraging and male punishment were observed over 10 trials per couple. Crucially, each trial consisted of two plate presentations P1 and P2, separated by a 60 s delay, allowing us to measure the effects of punishment on immediate future behaviour (as in [7]). During each plate presentation, we recorded the total number of flake items eaten by each cleaner, as well as the identity of the fish that ate the prawn item thereby causing the plate to be removed. In trials where females caused plate removal by eating the prawn item in P1, we recorded the intensity of male punishment (chases per second) during the 60 s interval. From these values, we calculated the mean intensity of male punishment in each experimental pair (\(n=21\)). Mean punishment intensity was square-root transformed (to achieve normality) and set as the response term in a linear mixed model (LMM) with normal error distribution and identity link function. Relative female size was included as an explanatory term, and female and male identities, respectively, were included as random terms to control for the non-independent effects of repeated measures taken from the same individuals. Female identity explained none of the variation in the data and was excluded from the final model.

We then asked whether female foraging behaviour changed in response to male punishment intensity by asking whether punishment intensity in response to female cheating in P1 affected the probability that females ate a prawn item again in P2. Data from 108 trials where females ate prawn in P1 were used. Female behaviour \((1=\text{ate prawn in P2}, 0=\text{did not eat prawn in P2})\) was set as the response term in a GLMM with binomial distribution of errors and a logit link function. Punishment intensity after P1, experimental treatment (‘similar-sized male’ or ‘dissimilar-sized male’) and the interaction between the terms were included as fixed terms. Female and male identities were included as random terms in the model to control for non-independence of repeated measures from the same individuals. Male identity was found to explain none of the variance in female foraging behaviour and was therefore excluded from the final model.

In both experiments, if the female hid in the shelter tube in the period following plate removal, then the number of chases was divided by the number of seconds that the female was out of the tube to obtain punishment intensity. Treatment order was counterbalanced within and between pairs. Experiments took place between 08.00 and 17.00 h and a minimum interval of 30 min (maximum = 60 min) was maintained between trials to prevent fish from becoming satiated. Fish were not fed between trials.

3. RESULTS

Male punishment of females was most severe, when females cheated high value clients and when the female was relatively close in size to the male (linear regression, \(F_{1,9}=10.4, p=0.01\); figure 1). However, males did not appear to significantly adjust punishment intensity according to client value when the female was relatively small in size (figure 1). Data from experiment 2 supported the idea that relative female size influenced male punishment intensity. Females that caused plate removal (by eating a prawn item) received more severe punishment from similar-sized males than from dissimilar-sized males (LMM, relative female size, likelihood ratio = 14.9, d.f. = 1, \(n=21\), \(p<0.001\), effect size \(\pm\) s.e. = 0.78 \(\pm\) 0.1; figure 2). Using data where females ate the prawn item during P1, we found that the more severe the punishment, the less likely females were to eat prawn again in P2 (GLMM, punishment intensity, \(\chi^2 = 8.70\), d.f. = 1, \(n=108\), \(p=0.003\), effect size \(\pm\) s.e. = -9.8 \(\pm\) 4.1; figure 3). This result controlled for the non-significant effect of experimental treatment (‘similar-sized’ or ‘dissimilar-sized’ male) (\(\chi^2 = 1.0\), d.f. = 1, \(n=108\), \(p=0.32\)).
Thus, females behaved more cooperatively in response to increased punishment from males.

4. DISCUSSION

Our findings show that males adjusted the intensity of punishment directed towards females according to the situation. Specifically, males punished females more severely when females both cheated a high value client and were close in size to the male partner. This adjustment was apparently adaptive since females behaved more cooperatively in response to increased punishment, in that they ate more against their preference. Although similar results have been demonstrated in humans [2–4,6], this is the first evidence for flexible adjustment of punishment and corresponding responses of targets in a non-human species.

Currently, the conditions for the evolution and the stability of punishment as a mechanism promoting cooperation, in particular in public goods games, are strongly debated both on the empirical level [22–26] and on the conceptual level [27–31]. The theoretical papers treat punishment and reaction as binary variables, while our empirical study and a few others [2,3] suggest that these variables are continuous. The assumption that investment in punishment is finely tuned according to the situation holds an intuitive appeal: in humans and non-humans alike, it is rare to observe invariant all-or-nothing behavioural responses to transgressions. Such differences may have a fundamental bearing on the use and efficacy of punishment by allowing players to tailor their investment in punishment according to the seriousness of the crime and the potential for future rewards based on changes in the punished individual’s behaviour. Ideally, these issues would be explored further by game theoreticians. Furthermore, although studies of punishment have typically assumed that all players are equal, with respect to their ability to punish and counter-punish one another [22–26,28] but see [32], in the light of empirical results presented here and previously [7], we suggest that player asymmetries may be an essential ingredient permitting the evolution of punishment to promote cooperation. Crucially, the assumption that all players are equal allows players to retaliate in response to punishment and this may act to destabilize rather than promote cooperation in social dilemmas (e.g. [22,33–37]). However, as first pointed out by Clutton-Brock & Parker [1], player asymmetries are a common

![Figure 2](image1.png)

**Figure 2.** The intensity of male punishment (square-root transformed) after females ate a prawn item according to the size asymmetry within the pair. The line was generated from the predictions of an LMM. Points are mean values of punishment intensity (square-root transformed) generated from raw data (n = 21 experimental pairs).

![Figure 3](image2.png)

**Figure 3.** The probability that females ate prawn again in the second presentation, P2 (after eating prawn in the first presentation, P1) according to male punishment intensity (chases per second, square-root transformed) after P1. The line was generated from the output of an LMM based on 108 trials, where females ate the prawn item in P1. Open circles are mean values generated from raw data (n = 21 experimental pairs).
feature of cooperative interactions [38,39]. Our results fit the ‘common-sense’ prediction that punishment is most likely to operate down a dominance hierarchy [1], with little scope for retaliation from subordinates. Indeed, it has been theoretically demonstrated that punishment may increase the average pay-offs to all members of a group if power asymmetries among players exist [32].

In conclusion, our results add to the growing literature that punishment may indeed promote cooperation in nature. We predict that both power asymmetries and the option for variable investment in punishment will turn out to be key conditions that make punishment more efficient than otherwise more suitable partner control mechanisms, such as cheating in return [22], partner switching [8] or premature termination of interactions (‘sanctions’ [40]). We also note that the scale of competition is likely to affect the conditions under which harmful behaviours, such as punishment, will evolve. Specifically, harmful behaviours towards social partners are expected to be more likely when the success of one individual negatively affects the fitness of its partner, owing to local competition for resources [41]. This idea applies to cleaner wrasse pairs where relatively large females represent a more immediate threat to male fitness owing to the increased likelihood of change sex compared with females that are relatively smaller in size.

N.R. and R.B. designed the study. N.R. collected and analysed data. R.B., S.W. and A.P. collected fish for experiments and helped with experimental design. N.R., A.G. and R.B. wrote the paper. All authors discussed results and commented on the manuscript.

The authors thank Lyle Vail and Anne Hoggett for support and assistance with data collection at Lizard Island Research Station (LIRS)—a facility of the Australian Museum. Thanks also to Sinéad English and Andrew Bateman for statistical help and to Joah Madden, Andy Gardner and two anonymous referees for useful comments. This work was funded by an LIRS John & Laurine Proud Fellowship (N.R.) and the Swiss National Science Foundation (R.B.).

REFERENCES


3 Fehr, E. & Gächter, S. 2002 Altruistic punishment in humans. Nature 415, 137–140. (doi:10.1038/415137a)


