

18 as well as cleaner and client fish densities from the locations of cap-
19 ture. Using Bayesian statistics to fit the model parameters to per-
20 formance data revealed that cleaner fish most likely estimate future
21 consequences of an action, while it appears unlikely that the removal
22 of the ephemeral reward acts as psychological punishment (negative
23 reinforcement). Incorporating future consequences also yields perfor-
24 mances that can be considered the result of locally optimal decision-
25 rules, in contrast to the negative reinforcement mechanism. We argue
26 that the combination of computational models with data is a powerful
27 tool to infer the mechanistic underpinnings of cognitive performance.

28 **Lay summary**

29 Cleaner fish eat ectoparasites off other fishes, so-called clients. It regularly
30 happens that two clients seek a cleaner’s service simulatenously. Cleaners
31 benefit from prioritising clients unwilling to wait, so they can feed on those
32 willing to wait. To make the right choice, cleaners must somehow “look”
33 into the future to anticipate consequences of current choices. By combining
34 a learning model with data, we show that cleaners estimate the long-term
35 value of their actions rather than using simpler heuristics. Estimating long-
36 term value is a mechanism involved in human foresight.

37 **Keywords**— learning, behavior, cleaners, bayesian statisitics, behavioral
38 mechanisms

39 Introduction

40 Often alternative cognitive mechanisms yield similar behavior and/or cogni-
41 tive performances. This poses a problem for disentangling the mechanistic
42 underpinnings of behavior. This is particularly clear in research aimed at
43 discovering between species variation in *higher* cognitive abilities; or in other
44 words, research on whether non-human animals show cognitive abilities be-
45 lieved to be uniquely human. For instance, when researchers try to find
46 *mental time travel-like* behavior, they usually come up with experiments to
47 show the behavior displayed requires inferences made through past events
48 (Dally, Emery, and Clayton 2006). However, they often face the challenge
49 of alternative scenarios where simpler explanations, like classic associative
50 learning, can bring about the observed behavioral outcome (Suddendorf and
51 Corballis 2007). Similarly, attempts to demonstrate the presence of *theory*
52 *of mind* in non-human animals face objections justified by alternative mech-
53 anisms underpinning similar behavioral results (C. M. Heyes 1998). Such
54 controversies are usually settled by using the principle of parsimony and its
55 cognitive version, Lloyd Morgan’s cannon, which states that the simpler ex-
56 planation (mechanism) should be accepted. Ideally, alternative hypotheses
57 should be evaluated in light of their explanatory power.

58 Learning is a key overarching cognitive mechanism that allows individuals
59 to associate rewards with environmental stimuli and thus behave adaptively
60 (Staddon 2016; Shettleworth 2009). Associative learning, in particular, exists

61 in all major vertebrate taxa, and in many invertebrates as well (C. Heyes 2012;
62 Macphail 1982; Staddon 2016; Behrens et al. 2008). Associative learning is
63 not homogeneous throughout its taxonomic distribution, rather there are
64 differences across and within species (Shettleworth 2009; Sih and Giudice
65 2012). So presumably, the mechanistic underpinnings of learning have been
66 modified by natural selection (Marler and Peters 1989).

67 One way to formalize the alternative mechanistic underpinnings of associa-
68 tive learning is to develop quantitative models of learning processes. This ap-
69 proach, which started within experimental psychology (Staddon 2016; Bran-
70 don, Vogel, and Wagner 2002), has been very fruitful in disentangling the
71 mechanistic structure of cognitive systems. More recently, the development
72 of reinforcement learning theory (Sutton and Barto 2018), has allowed to
73 evaluate the empirical support for alternative mechanistic hypotheses by
74 providing quantitative predictions which are amenable to statistical tests
75 (Farashahi et al. 2020, 2017). Interestingly, these learning models not only
76 have received support from behavioral data, but also are consistent with the
77 current view on reward processing in the brain (Schultz 2015).

78 From an evolutionary perspective, mechanisms are likely selected because
79 of how they allow individuals to respond to environmental variation. For
80 example, biological market theory predicts that the exchange rate of goods
81 and/or services traded between cooperative partners adjusts to the law of sup-
82 ply and demand, when individuals have some degree of partner choice (Noë

83 and Hammerstein 1995). Supply and demand conditions, which typically
84 depend on the abundance of the species involved, certainly vary in time and
85 space. Therefore, natural selection should favor the ability to flexibly adjust
86 decisions and behavioral output to current market conditions. Indeed, such
87 adjustments have been documented (Axén, Leimar, and Hoffman 1996). One
88 example of strategic adjustment in a biological market is the marine cleaning
89 mutualism involving the cleaner fish *Labroides dimidiatus* and ‘client’ fishes.
90 Client fishes seek cleaner fish services at their territory (so-called “cleaning
91 station”) and offer themselves as food patches to get their ectoparasites re-
92 moved, which provides cleaners with food and clients with improved health
93 (Waldie et al. 2011; Ros et al. 2011; Triki et al. 2016; Demairé et al. 2020).
94 Given the capacity of some client fish to swim larger distances and access mul-
95 tiple cleaning stations while others access the only cleaning station in their
96 territory, it is crucial to categorize clients as either “visitors” or “residents”,
97 respectively. During cleaning interactions, a cleaner fish often faces a choice
98 between a visitor and a resident client seeking its cleaning services simultane-
99 ously. Visitors have the option to switch to another cleaner fish if being made
100 to wait, while residents must wait for inspection. Indeed, visitors have been
101 observed to use their partner choice option in that way (Bshary and Schäffer
102 2002), which may explain why cleaners give visitors service priority in a field
103 study in the Red Sea (Bshary 2001). Furthermore, in a lab based paradigm,
104 design to mimic the resident-visitor choice (ephemeral reward task), cleaners
105 learned to prefer the cue associated with the epheral food source (visitor)

106 and hence accessed both food sources, obtaining double the amount of food
107 (Bshary and Grutter 2002). However, further exploration revealed that not
108 all cleaner fish manage to develop a preference for the ephemeral option in
109 the lab (Triki et al. 2018). Over the last decade, over a hundred wild-caught
110 cleaner fish have been tested in the exact same paradigm of the ephemeral
111 reward task (Salwiczek et al. 2012; Wismer et al. 2014; Triki et al. 2018,
112 2019, 2020). These fish often come from different reef locations. Investigation
113 of the local eco-sociological conditions revealed that cleaner and client fish
114 population densities have a substantial impact on cleaner fish performance
115 in the task. Cleaner fish from reef sites with relatively low densities were
116 more likely to fail at solving the task (Triki et al. 2019, 2018; Wismer et
117 al. 2014). This intra-specific variation is unlikely due to local genetic adap-
118 tation, because cleaner fish are open water spawners and the environmental
119 conditions can vary within the lifespan of a fish.

120 Mechanistic models explicitly designed to mimic the ephemeral reward task
121 have shown that the simplest form of associative learning (operant condi-
122 tioning) cannot account for a solution to the ephemeral reward task (Prat,
123 Bshary, and Lotem 2022; Quiñones et al. 2019). Operant conditioning is
124 a form of associative learning where individuals use short term reward to
125 associate and choose actions. Such models allow varying the cognitive tool
126 kit and evaluating which minimal kit is necessary to solve the task at hand
127 (e.g. Dubois et al. (2021)). To be able to give visitors priority over resi-
128 dents, cleaners need to be able to assess a client’s value separately for the

three possible scenarios (alone, paired with a fish with the same strategic option, paired with a fish with the alternative strategic option) (Quiñones et al. 2019). The ability to distinguish and value one stimulus differently alone from compound versions of it has been termed configurational learning, chunking, or segmentation (see references in Prat, Bshary, and Lotem (2022)). In addition to configurational learning, cleaners also need to account for the future consequences of current decisions. In the model by Quiñones et al. (2019), this could be achieved in two non-mutually exclusive ways: through low temporal discounting of future effects, also termed ‘chaining’ (Enquist, Lind, and Ghirlanda 2016); and/or through perceiving a visitor client leaving as psychological punishment (i.e. as a negative reinforcer). Chaining is when individuals include in their valuation of an action the reward effects that this will have in the future. This is done by combining in a single valuation the reward obtained in the current time with all the reward that comes after, discounting for how far in the future reward is accrued. ‘Chaining’ the reward of these different time steps allows individuals to take actions that increase the long-term reward at the sacrifice of short term considerations (Enquist, Lind, and Ghirlanda 2016). Even though, ‘chaining’ can be readily implemented computationally in learning models (Enquist, Lind, and Ghirlanda 2016; Sutton and Barto 2018), cognitively it seems to be a complex adaptation (Suddendorf and Corballis 2007). On the other hand, using client behavior as a negative reinforcer is, in principle, easier to implement. Thus, the standard logic of Lloyd Morgan’s cannon demands that operant

conditioning as the simpler explanation is to be accepted by default. Ideally, however, the two mechanisms should be evaluated in light of how well they explain the available data. Note that different fields interested in cognition and decision making use different words to refer to negative reinforcers (Quiñones et al. 2019; Sutton and Barto 2018). Here, for the sake of simplicity and clarity, we will use the word ‘penalty’ to refer to this mechanism which includes a negative reinforcer.

In here we used the field and experimental data to fit the parameters of a reinforcement learning model to infer the cognitive mechanism that cleaners use in their interaction with clients. Specifically, our approach of fitting the computational model to the empirical data aimed at: (i), determining which mechanism cleaner fish use to incorporate future consequences of current decisions by testing whether chaining, penalty, or a combination of both best explains their performance; (ii) determining whether the two mechanisms differed with respect to the ecological conditions that are likely to cause high versus low performance in the ephemeral reward task. Additionally, we assessed which mechanism yields optimal performance patterns. Relying on the logic of biological market theory, we predicted that appropriate performance is to show a high preference for visitors only under high local cleaner-to-client ratio.

172 **Methods**

173 **The model**

174 The model consists of a set of individual-based simulations where individuals,
175 representing cleaner fish, face a series of choices between two options, which
176 simulate the natural conditions of the cleaning market. Individuals experi-
177 ence a series of discrete time points in which they face different ‘states’, de-
178 fined by the number and category of client fish (visitor or residents) inviting
179 for cleaning services. There are six possible states: zero clients, one resi-
180 dent, one visitor, resident-resident, visitor-visitor, and resident-visitor. The
181 probability of each state is largely determined by the relative abundance of
182 cleaner fish, residents and visitors, but to some degree by cleaner fish choices
183 when it faces the resident-visitor combination. This is because residents are
184 willing to queue for cleaning service; while visitors leave the queue (with a
185 certain probability) when made to wait. Individuals obtain a fixed reward
186 from cleaning a client fish regardless of the category. Every time individuals
187 face and make a choice they update the probability of making that same
188 choice. The update is based on the difference between the expected value
189 and the obtained reward - the prediction error (δ_t) - (Sutton and Barto 2018;
190 Rescorla and Wagner 1972). Formally, the prediction error is given by

$$\delta_t = R_t - V_t(S_t) + \gamma V_t(S_{t+1}), \quad (1)$$

191 where R_t is the sum different reward sources at time t ; $V_t(S_t)$ is the estimated
 192 value at time t of the the state faced by the agent at time t ; similarly $V_t(S_{t+1})$
 193 is the estimated value of the state to come in the following time-step, γ is the
 194 discount factor for future rewards. When the estimated value of the current
 195 state ($V_t(S_t)$) is equal to the sum of short-term (R_t) and future discounted
 196 reward ($\gamma V_t(S_{t+1})$) learning stops for that state. If $\gamma = 0$ the estimates
 197 made by the agent only capture short-term reward. We assume short-term
 198 reward to two components: positive reward determined by the amount of
 199 food obtained from cleaning a client; and negative reward triggered when
 200 by a client leaving the station without being cleaned. Formally, we let total
 201 reward be given by $R_t = P_t - \eta_t$. Where η is a parameter of the model
 202 that determines the the size of the negative reward triggered by unattended
 203 clients leaving the station.

204 The prediction error (Eq. 1) is used to update the value of each one of the
 205 states the agent faces, as well as the preference for the resident and visitor
 206 options. The value update is simple the product of the prediction error
 207 and the parameter for the speed of learning ($\Delta V(S_t) = \alpha \delta_t$). The change
 208 in preference between the resident and visitor is given by $\Delta(\theta_v - \theta_r)_t =$
 209 $\alpha \delta_t 2(1 - \pi_v)$, where θ_i represent the preference for one of two options and the
 210 difference captures the total change relative to one another; π_v corresponds to
 211 the current probability of choosing the visitor. p_i is determined by applying
 212 the logistic function to the diffence in preferences between the two mutually
 213 exclusive options ($\pi_v = \frac{1}{1+e^{-(\theta_v-\theta_r)}}$). This amounts to a preference update

214 that is carried in the direction that leads to more reward being obtained,
215 given the new information. In the long run, the probability of choosing a
216 visitor over a resident converges in the model. To which probability the
217 model will converge depends on the relative abundance of cleaners, visitors
218 and residents; as well as on the probability of visitors leaving the cleaning
219 station when made to wait. Further details of the model implementation can
220 be found in Quiñones et al. (2019).

221 The model shows that agents need to find a way to incorporate future conse-
222 quences of current choices. In the model, this could be achieved with either of
223 two parameters that could also work together. First, γ measures how much
224 individuals include future rewards in their decision updates. If $\gamma = 0$, indi-
225 viduals only use the immediate reward obtained from a cleaning interaction.
226 As γ increases, individuals include more the reward obtained from the sub-
227 sequent choices. That amounts to estimating and using for decision making
228 the future expected rewards of an action (chaining). Second, η measures how
229 much individuals include in their reward the fleeing behavior of visitors as
230 a negative component (penalty). Both of these parameters allow individuals
231 to use in their estimates the future effects of their choices.

232 **Empirical data**

233 The empirical data were collected between 2010 and 2019 always during the
234 austral winter months June to August from a total of five study reef sites

235 (Corner Beach-CB, Horseshoe-HS, Mermaid Cove-MC, Northern Horseshoe-
 236 NHS, and The Crest-TC) at Lizard Island ($14.6682^{\circ}S, 145.4604^{\circ}E$), Great
 237 Barrier Reef, Australia. The data consist of three sets: fish censuses, field
 238 observations of cleaner-client interactions to quantify the probability of vis-
 239 itors leaving if made to wait, and the performance of wild-caught cleaner
 240 fish in the ephemeral reward test. In total, we have twelve site/year data
 241 sets for fish censuses and corresponding performance in the lab test. Thus,
 242 some sites were sampled more than once. To estimate the population density
 243 of cleaner fish and their clients at a given site in a given year, Triki et al.
 244 (2019) used a series of ten transects of $30m$ each. Observers swam along the
 245 transect lines placed on the reef and first counted the visible large-bodied
 246 adult fish (species with total length $TL \geq 10cm$) including cleaner fish on a
 247 width of $5m$, and then on the return individuals of small-bodied fish species
 248 ($TL < 10cm$) on a width of $1m$ (see Triki et al. (2018) for further details
 249 on fish censuses data collection). We then scaled the counts of cleaner fish,
 250 small-bodied, and large-bodied clients fish densities per $100m^2$. From the
 251 study by Triki et al. (2019) cleaner fish and large-bodied client populations
 252 densities were highly correlated, and only the former was hence used in the
 253 analyses as representative of both measures.

254 The field observation data consisted of video recordings/encodings of the
 255 cleaner-client cleaning interactions. There were videos from eight cleaners
 256 per site/year of a duration of $30min$ each. Triki et al. extracted information
 257 from every event wherein a visitor client was made to wait in favour of another

client (visitor or resident), and noted whether or not the visitor left or queued for the cleaning service (Triki et al. 2019, 2020).

The cognitive performance data was from a total of 120 cleaners (10 individuals per 12 site/year) tested in the ephemeral reward task (Triki et al. 2019, 2020). Authors housed all captured cleaners individually in glass aquaria ($62\text{cm} \times 27\text{cm} \times 37\text{cm}$) and provided them with PVC pipes ($10\text{cm} \times 1\text{cm}$) as shelters. The task consisted of exposing the cleaner fish to substitute models of client fish in the form of two *Plexiglas* plates offering the same amount of food (one item of mashed prawn). The two plates differed in colour and pattern (horizontal green stripes or vertical pink stripes) but had equal size ($10\text{cm} \times 7\text{cm}$). Importantly, the two plates played different roles as either a visitor (ephemeral food source) or resident (permanent food source). That is, if a cleaner fish inspected the resident plate first, the experimenter withdrew the visitor plate out of the aquarium as a consequence. Choosing first the visitor plate, however, granted access to both plates. The equal size of the plates forced cleaner fish to learn to give service priority to the visitor plate based solely on the behavior-cue of the plates rather than size-cue (Wismer et al. 2019). Triki et al. (2019, 2020) tested the fish for a maximum of 200 trials with 20 trials a day, 10 trials in the morning and 10 trials in the afternoon. They randomized and counterbalanced the plates' spatial location (i.e. left or right) between trials. Similarly, they counterbalanced the plates' decoration (colour and pattern) and the plates' role (visitor or resident) between the tested fish. In the original studies, once a fish reach a learning criterion, that

281 is, performing significantly above chance level (> 50 , $p - value \geq 0.05$), they
 282 passed to a reversal version of the task where the roles of the visitor/resident
 283 Plexiglas plates were swapped (see Triki et al.(2019, 2020)). Here, we used in-
 284 stead a subset of these data in order to have an idea of cleaner fish preferences
 285 for the visitor plate, even if they do not reach the learning criterion within
 286 200 trials. To do so, we first extracted the trial-by-trial outcomes from the
 287 last two sessions (20 trials) of those who never reached the learning criterion
 288 for visitor plate ($N = 45$ cleaner fish). For those who reached the learning
 289 criterion at some point during the test and passed to a reversal phase, we
 290 extracted the trial-by-trial outcomes from the last session (10 trials) before
 291 passing to reversal and the last 10 trials they were exposed to in the test
 292 ($N = 75$ cleaner fish). We chose a combination of initial and reversal to
 293 quantify preference for the visitor client, given that it matches well the cri-
 294 teria chosen in previous analysis of the ephemeral reward task experimental
 295 set-up (Triki et al. 2019, 2020). In the supplementary material (Fig. S1) we
 296 show how our choice maps to the previously used criteria. For comparison,
 297 we also show a quantification of visitor preference based only on the initial
 298 phase of the task.

299 **Statistical analysis**

300 The aim of the analyses is to fit the key model parameters γ and η , to the
 301 empirical data from Triki et al. (2019, 2020) to test whether each or a

302 combination of these effects is a better explanation for the pattern seen in
303 the data. We used the ecological variables: cleaners, visitor clients, resident
304 clients abundances and visitor clients leaving probability,
305 as input to the models. As the response variable, we used cleaners' preference
306 for visitors over residents in the ephemeral reward task. Finally, we used the
307 preference for visitors resulting from the model simulations as the prediction
308 for the response variable. We kept all other parameter values used for the
309 model simulations constant, see Table S1.

310 To capture with the model the relationship between the ecological variables
311 and cleaner fish preferences for visitors, we needed to scale the absolute popu-
312 lation densities of cleaner fish from the empirical data to a measure of relative
313 abundances that captures client visitation patterns. This is because, in the
314 model, relative abundances of clients define not only the probability of resi-
315 dents and visitors but also how often the cleaning station is empty (e.g. there
316 are no clients to be cleaned). The frequency with which clients visit the sta-
317 tion is another variable influencing the station occupancy, which in nature
318 may vary among different client species depending on their ectoparasite loads.
319 We do not have field estimates for species-specific parasite loads, especially
320 not as a function of the site. In order to control for these aspects, we com-
321 puted a measure of relative cleaner fish abundance for each reef site relative
322 to absolute abundances and multiplied it by a scaling constant that changes
323 the range of the variable. This scaling constant is hence meant to capture
324 variation in the market conditions driven partly by cleaner fish abundance.

325 We fitted the value of the scaling constant as part of the statistical inference.
 326 As for the visitor and resident abundances, we computed a relative measure
 327 with respect to the total client abundance, and weighted that by the rescaled
 328 relative absence of cleaners ($1 - \text{relative cleaner abundance}$). Thus, all three
 329 measures of relative abundance sum up to 1, and can be used in the model
 330 as a proxy for the probability of having different options in the cleaning sta-
 331 tion. Note that we introduced the scaling constant to control for variation
 332 that is not captured by the model; its parameter distribution does not offer
 333 biological insights.

334 Once we calculated the relative abundances, we obtained predictions from the
 335 model for each one of the locations and ran the Markov Chain. We started
 336 the chain with random values for the parameters of interest, then ran the
 337 computational model once for every reef site using as input the ecological
 338 explanatory variables. The model outputs the probability of choosing the
 339 visitor for each location p_i , where i is the index for the 12 locations. Assum-
 340 ing a binomial distribution, the probability that each of the cleaner fish 20
 341 choices in the ephemeral reward task was generated by the model is given
 342 by $\binom{n}{k_{ij}} p_i^{k_{ij}} (1 - p_i)^{n - k_{ij}}$; where k_{ij} is the number of times that cleaner fish
 343 j in reef site i chose the visitor over the resident; and $n = 20$ (due to our
 344 choice of using 20 choices per cleaner fish). By taking the natural logarithm
 345 and summing over all the cleaner fish and reef sites, we obtained the log-
 346 likelihood of the data given the model and parameters. We then proposed
 347 a new set of parameters drawn from a uniform distribution centred around

348 the old parameter set. The amplitude of the uniform distribution used for
 349 each parameter can be found in Table S1. Subsequently, we ran the model
 350 and calculated the likelihood with the new parameter set. We then used
 351 the ratio of the two likelihoods to choose which parameter set to keep. New
 352 parameter sets with a higher likelihood than the old set replaced old ones,
 353 and those with a lower likelihood replaced current ones with a probability
 354 equal to the log-likelihood ratio. Given that we only used the likelihoods
 355 in the decision, we used an uninformative prior. Once we decided whether
 356 the new parameter set would replace the old one, we ran the model again to
 357 sample the likelihood distribution of the parameter set. We then started the
 358 cycle again by proposing a new set of parameters and repeated the process
 359 for $1e^5$ steps. We ran 5 independent chains, discarded the first 1000 samples
 360 of each chain as burn in, and after that, we kept 1 in every 100 samples to
 361 avoid autocorrelation. The collection of parameters sets kept in all chains
 362 approximates the posterior distribution. We coded the model as well as the
 363 fitting algorithm in *c++*, the diagnostics and visualization in R (R Core
 364 Team 2021). All codes are accessible at <DOI: 10.5281/zenodo.6338073>

365 To compare the fit of the three alternative models, we used the distribution of
 366 pseudo R^2 proposed by Mcfadden (McFadden 1974). Mcfadden’s pseudo R^2
 367 is a standard measure of fit for logistic regression. In that context, *pseudo*—
 368 R^2 uses the log-likelihood of the data given the model, relative to the log-
 369 likelihood of the data given a model without covariates, as a measure of fit.
 370 Our model is not a logistic regression, therefore we measured the *pseudo*— R^2

371 in relation to the log-likelihood of a model with parameters γ and η set to
 372 zero. This, in practice, amounts to a model that has a neutral preference
 373 between the two options. We computed the pseudo R^2 for all the samples of
 374 the posterior from the MCMC. We used these distributions of pseudo R^2 's
 375 as a measure of fit.

376 **Results**

377 Estimation of parameter values for the three models (full model, chaining and
 378 penalty), support chaining as the only mechanism cleaners use to account for
 379 the future effects of their actions; and thus to solve the ephemeral reward task.
 380 In the estimation of the parameter values of the full model, which includes
 381 both chaining and penalty, the bulk of the marginal posterior distribution
 382 of η which controls the strength of penalty is around 0 (Fig. 1). As for γ ,
 383 controlling chaining, the 95% confidence intervals also includes zero, but the
 384 mode of the posterior is around 0.5 (Fig. 1, a). In the chaining model, where
 385 η is set to zero, the distribution of γ shifted to higher values, zero is no longer
 386 part of the 95% credible interval of the parameter (Fig. 1, c). In contrast,
 387 when we look at the model with only penalty, the posterior distribution of η
 388 is still centred around zero (Fig. 1, d). Thus, the analysis of the estimates
 389 of individual parameter values in the three models only supports a strong
 390 effect of chaining. Furthermore, the comparison of the models' fit favours the
 391 chaining model. In panel e of figure 1 we show the distribution of *pseudo- R^2*

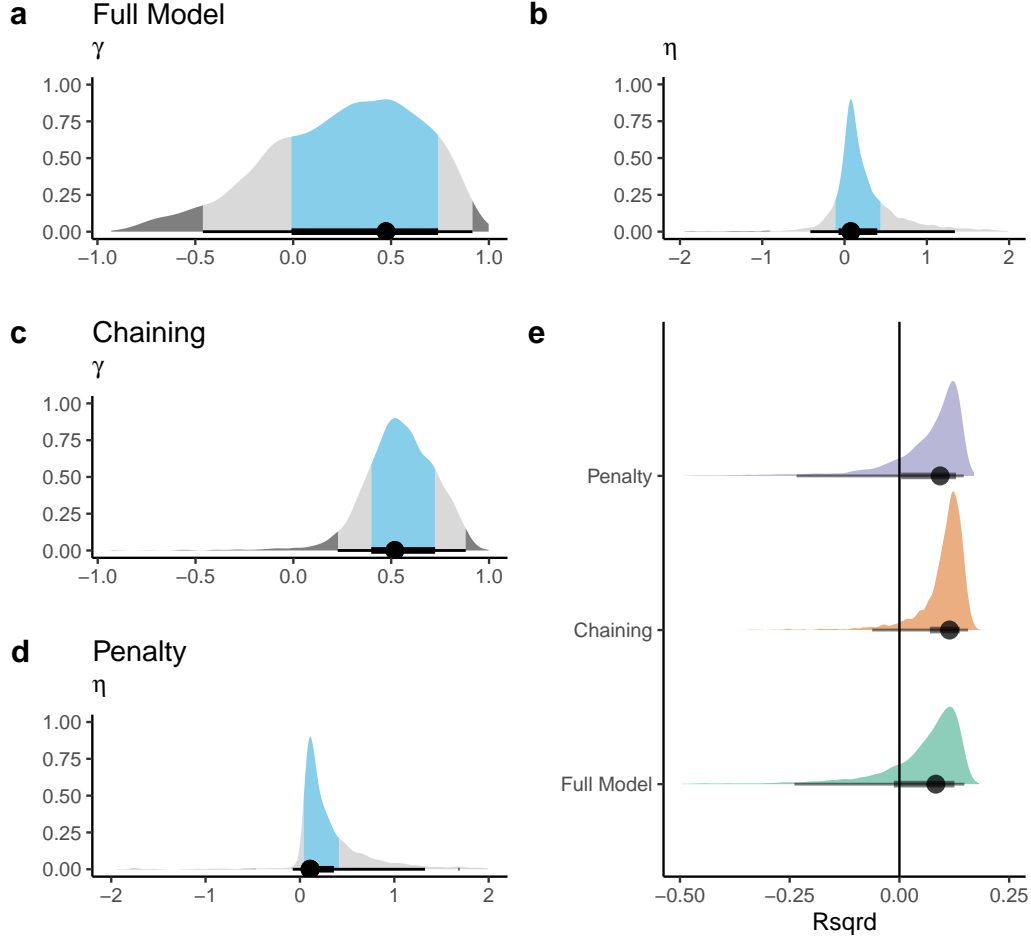


Figure 1: Posterior distributions for parameter values γ and η for the three models (Full model, chaining only and penalty only). We show the kernel density estimates, below the mode (black dot) and the 66 (light blue shade) and 95% (grey shade) highest posterior density interval for the two parameters. On the top, panels a and b show posterior distributions for a full model, including chaining (γ) and penalty (η). Panel c, shows the γ estimate from model with only chaining. Panel d shows the η estimate from a model with only penalty. Panel e, shows a measure of fit for all models, namely the distribution of pseudo- R^2 obtained from sampling the posterior distribution of parameter values.

392 calculated using samples from the posterior distributions shown before. Note,
 393 $pseudo-R^2$ can have negative values, which is when the log-likelihood of the
 394 model is lower than that of a model that triggers neutral preferences. Even
 395 though the peak of the three $pseudo-R^2$ distributions were not very different,
 396 the model with only chaining produced a distribution of $pseudo-R^2$ where
 397 more values were positive (to the right of the black line in Fig. 1 e). This
 398 shows that accounting for variation in the parameter estimates the model
 399 with chaining gives a better fit to the data, despite having one parameter
 400 less than the full model. We have not shown here the marginal posterior
 401 distributions of the scaling constant, given that they do not bring biological
 402 insight. Their visualizations can be found in the supplementary material
 403 (Fig. S2), as well as the diagnostics of the MCMCs (Figs. S3,S4,S5).

404 The main reason for chaining and penalty to give different predictions is the
 405 way that cleaner fish relative abundance influences the preference for the
 406 visitor clients. Visitor leaving probability has a similar positive effect on the
 407 probability of choosing the visitor clients on all three models, they all predict
 408 an increase preference for visitors as the visitor probability increases (Fig. 2).
 409 In contrast, cleaner fish relative abundance has a different effect in the model
 410 with only chaining, compared to the other two models, full model and penalty.
 411 In the model with only chaining, only intermediate cleaner fish abundance
 412 triggers a preference for the visitor clients (Fig. 2 c). In both models with
 413 penalty, both intermediate and low cleaner abundance triggered a preference
 414 for the visitor (Fig. 2 a,e). Note, however, we calculated preferences shown in

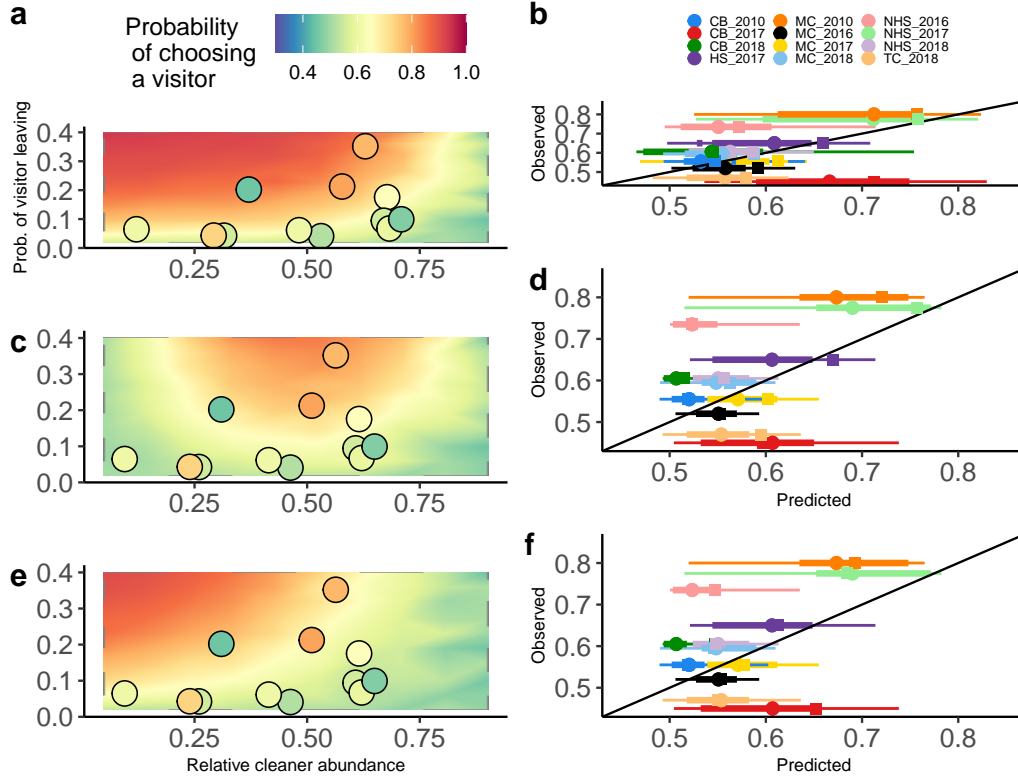


Figure 2: Observed and predicted probability of choosing a visitor. Left-hand side panel: colour contour shows the prediction from the learning model using the mode of the posterior distributions of parameters recovered by the statistical analysis. Dots show the frequency of visitor choices for the 12 reef sites, as well as the corresponding relative cleaner fish abundance (x axis) and frequency of visitors clients leaving the cleaning station (y axis). Right-hand side panels: Variation of the predicted probabilities of choosing a visitor over a resident and their observed values for 12 locations. Circles show the mean prediction for each location from 100 samples taken from the posterior distribution. Thick and thin bars show the 66 and 95% credible interval, respectively, taken from those posterior samples. Squares show the predictions used for the panel on the right-hand side. Colour coding denotes different reef site/year of the data collection (see Empirical data section). The black line corresponds to a perfect match between observed and predicted probabilities. Upper panels (a and b) show predictions from a model including chaining and penalty; middle panels (c and d) from a model with only chaining; lower panels (e and f) from a model with penalty only.

figure 2 left panels by using only the mode of the posterior distributions, and by holding constant the balance between resident and visitors' abundances. Panels on the right, show how close predictions are from the observed data, allowing the balance between client types to vary and using a set of samples from the posterior distribution.

Discussion

In this study, our main aim was to unravel which of two potential cognitive mechanisms, chaining of events, penalty, or their combination, best explains wild-caught cleaner fish performance in the ephemeral reward task, while accounting for their ecological conditions. To evaluate the merits of each of these two mechanisms separately and combined, we considered cleaner fish performance in the lab test to have its origin from the rule these fish applied in their natural environment. That is, individuals that solved the task already had a preference for visitor clients and generalized this rule to the lab conditions once being familiar with the task.

While all three models captured well the positive relationship between visitor leaving behavior and cleaner fish performance in the market task (Triki et al. 2019), only the chaining mechanism predicted that cleaner fish performance in the task should be low in habitats with low cleaner-to-client ratios, regardless of the visitor leaving probability. In contrast, models including negative reward predicted the highest performance in the ephemeral reward

task when relative cleaner fish abundance is low, particularly together with
a high probability of visitor leaving (Fig. 2). Low relative cleaner fish abundances mean the market has an excess of demand for cleaning services. In the models, this translates to a cleaning station that is frequently full. Thus, when a visitor leaves, it is likely that the cleaner fish will have access to another client in the next step. Therefore, there will not be much difference in future reward between choosing a visitor and a resident, and cleaners will not develop a preference for the visitor in these conditions. On the other hand, the effect of negative reward on cleaner fish preference is the opposite, as in a busy cleaning station, to that of chaining. Cleaner fish will get more often the resident-visitor state and will develop a preference for the visitor faster. At high cleaner fish abundances, the resident-visitor state becomes so rare that neither mechanism is very efficient at generating a preference for visitors. When facing the resident-visitor choice, it is still best to choose the visitor; however, the learning machinery will not be able to develop this preference efficiently. Overall, the models suggest that chaining is the cognitive mechanism that allows cleaner fish to adaptively adjust to their biological market ecological conditions.

Previous research showed that cleaner fish living at high population densities and giving service priority to the visitor plate in the ephemeral reward task, as well as cleaner fish living at low densities but denying service priority to the visitor plate possess larger forebrains; a key teleost brain region associated with behavioural flexibility and social intelligence. Those failing

459 to show optimized decision-rules given their local ecological conditions had
460 relatively smaller forebrains (Triki et al. 2020). Triki et al. refer to the for-
461 mer as socially competent cleaner fish, while the second group as socially
462 incompetent cleaner fish. Social competence is the ability to optimise social
463 behavior to the available social information (Taborsky and Oliveira 2012;
464 Bshary and Oliveira 2015; Varela, Teles, and Oliveira 2020). Our analyses
465 yielded no evidence that the difference in social competence with respect
466 to the local ecological conditions and associated brain morphology, found
467 by Triki et al. (2020), is due to the mechanism used to incorporate future
468 consequences. It is conceivable that high performing individuals from low
469 population densities reef sites use negative reinforcement instead of chaining,
470 but in that case, negative reinforcement should have explained at least part
471 of the data. Configurational learning or chunking (Sutherland and Rudy
472 1989; Miller 1956), the second component necessary to solve the ephemeral
473 reward task (Quiñones et al. 2019), was not varied in the models we analysed
474 here. However, while chunking tendencies should vary to allow individuals
475 to adapt to their local conditions (Prat, Bshary, and Lotem 2022; Kolodny,
476 Edelman, and Lotem 2014), systematic differences in individual chunking
477 tendencies would not explain how socially competent decisions vary as a
478 function of relative abundance. Therefore, it remains currently unclear what
479 cell-demanding mechanisms may cause variation in social competence that
480 translates into site-specific variation in performance in the ephemeral reward
481 task.

Our models are inspired by the general processes of associative learning where short term rewards are translated into decision making; thus, it ignores alternative channels of information that could be relevant in market-like situations. For example, the model does not investigate whether cleaner fish actually assess the frequency of client visits or a mean frequency of visitors leaving. The updating learning mechanism for the development of preferences works on a trial-by-trial basis. In the model, cleaner fish do not need to assess the actual state of the market, *i.e.* their abundance, the abundance of residents and visitors, and client visitation rate as an indicator of demand. They only need to assess the short-term consequences of their own decisions on food intake and chain them. Also, for the sake of simplicity, the model ignores the process by which cleaner fish discriminate residents and visitor clients. A model that accounts for this discrimination probably would involve the development of preferences for morphological or behavioural features that are statistically associated with visitors or residents. For example, visitors are on average larger than residents in body size (Bshary 2001), and contrary to residents, they are less likely to chase a cleaner fish that fails to cooperate and instead cheats its client by taking a bite of mucus (Bshary and Grutter 2002). Given these associations, chaining might produce the decision-rule “choose the larger client and/or the less aggressive client”, which is not a useful rule in the standard ephemeral reward task.

In conclusion, our study shows that variation in cognitive performance as a function of the local ecological conditions may set the stage for the use of

mechanistic modelling to identify the cognitive processes underlying learning in animals. The combination overcomes the limitations of the general philosophy in animal cognition to apply the logic of Lloyd Morgan’s canon (Occam’s razor). Cognitive experiments with the aim of excluding basic reinforcement learning as a potential explanation (operant and/or classical conditioning) of performance often employ one trial experiments requiring animals to solve the task on the first possible occasion. For example, any theory of mind task needs to be solved in the first trial in order to exclude fast conditioning (C. M. Heyes 1998). Similarly, subjects need to solve a social learning task on the first trial to accept imitation as a mechanism over stimulus/local enhancement. Such strict conditions are virtually never met. For example, potato washing by Japanese macaques, an iconic example of social learning, took several years to spread within the group (Kawamura 1959), meaning that any learner had been repeatedly exposed to demonstrations before acquisition. Importantly, Galef (1992) refuted imitation as a mechanism not simply because of the repeated exposure but because a (rather qualitative) analysis of the spread of potato washing across individuals did not follow the prediction based on imitation learning (see also Hirata, Watanabe, and Masao (2001)). In our case, the number of trials it took cleaners to learn the solution to the ephemeral rewards task would never allow excluding an important role of penalty based on the data alone. However, fitting model predictions to our comprehensive empirical data set revealed that a more complex mechanism, estimation of future reward, fits the data better.

Supplementary material

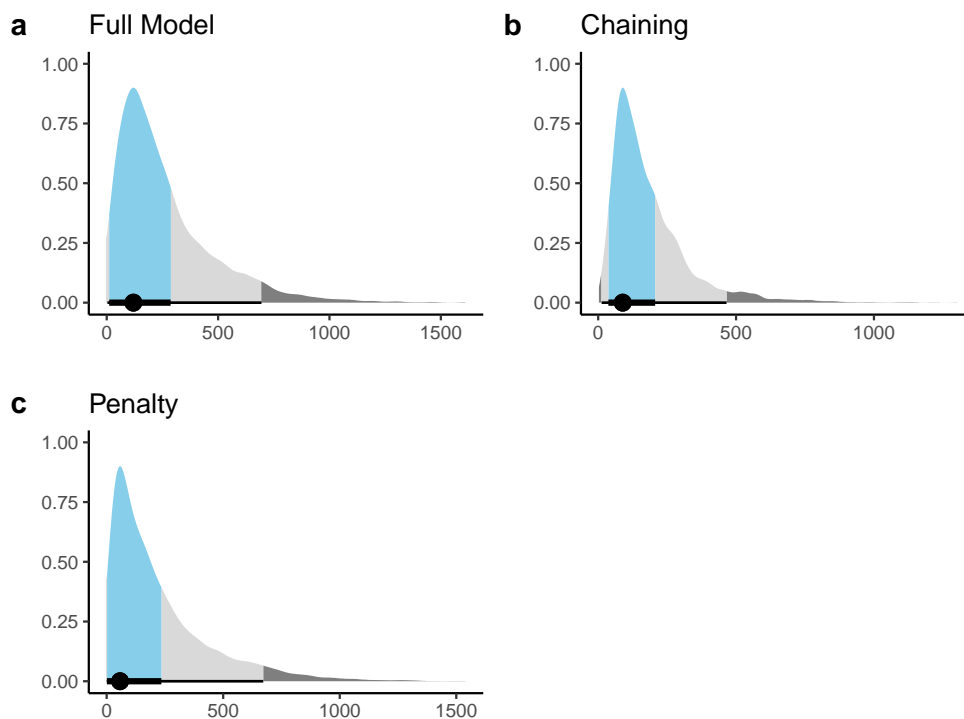


Figure S2: Posterior distributions for scaling constant for the three models (Full model, chaining only and penalty only). We show the kernel density estimates, below the mode (black dot) and the 65% (light blue shade) and 95% (grey shade) highest posterior density interval. On the top, panel a shows the posterior distribution from the full model; panel b from the model with only chaining; and panel c from a model with only penalty.

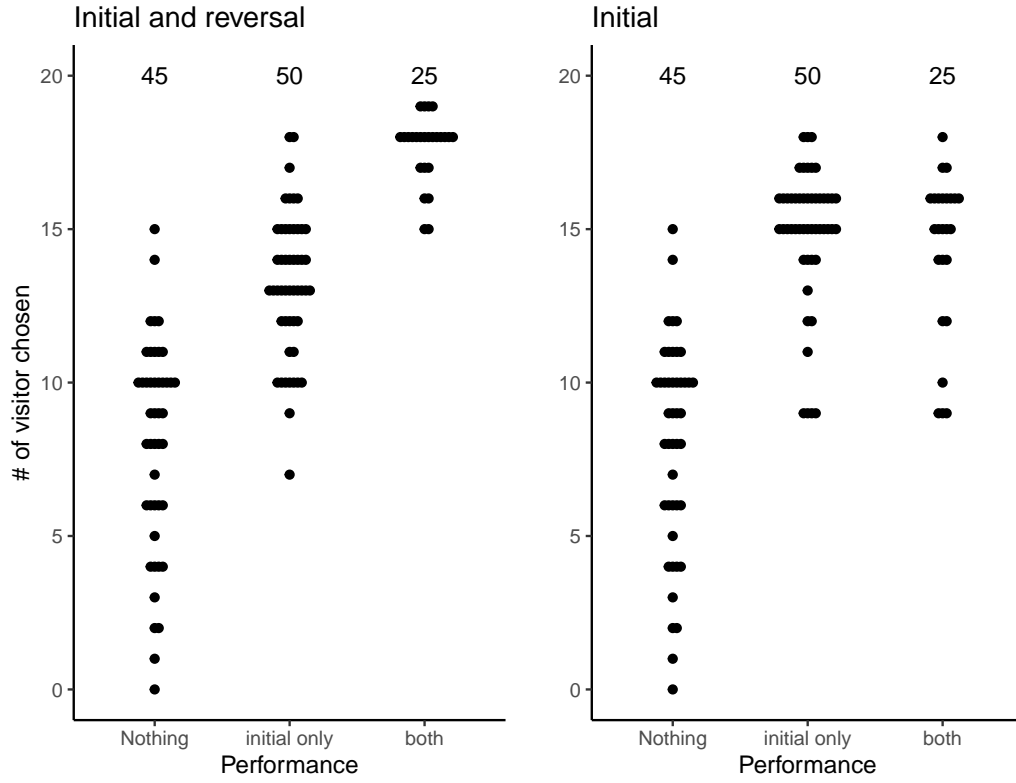


Figure S1: Relation between the response variable used in this study and the criteria used in previous studies to assess performance in the ephemeral reward task. In the x axis, we classified the performance of cleaner fish according to whether they developed a preference for the visitor in the initial round, in the initial and reversal, or none of them. In the y axis, we add the choices of two experimental sessions: panel on the left uses one session from the initial round and one from the reversal round when possible (as described in the main text); panel on the right uses two sessions from the initial round for all fish.

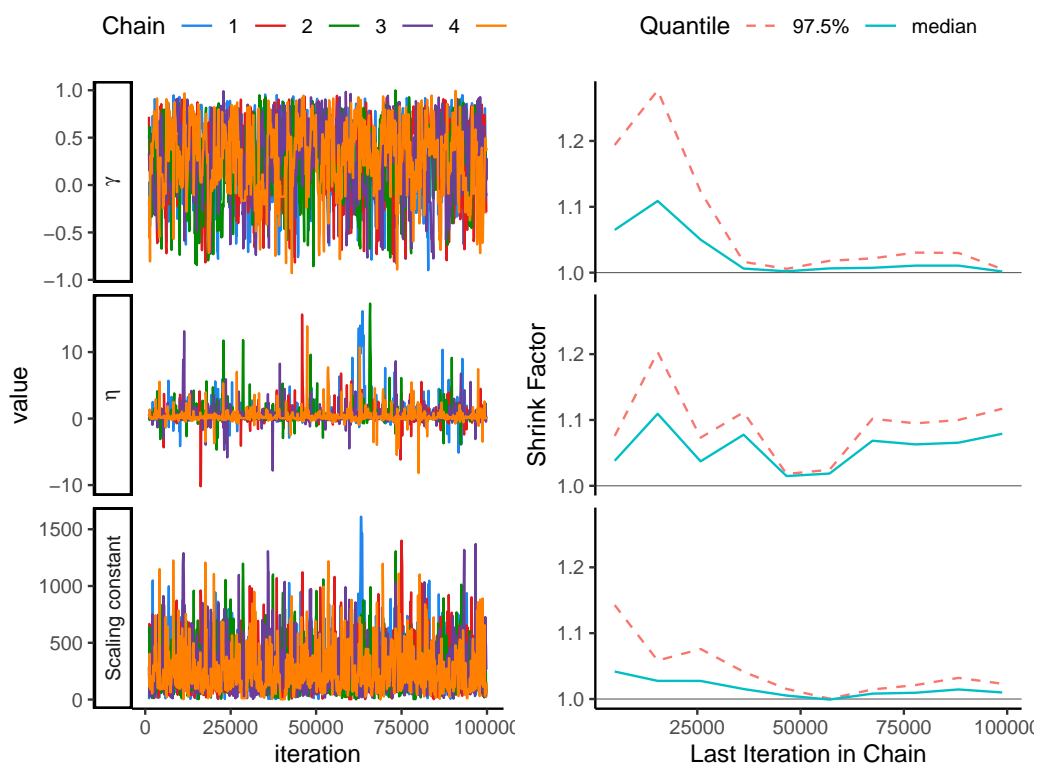


Figure S3: MCMC convergence diagnostics for the full model. On the left trace-plots, on the right changes along the chain of the Gelman and Rubin shrink factor.

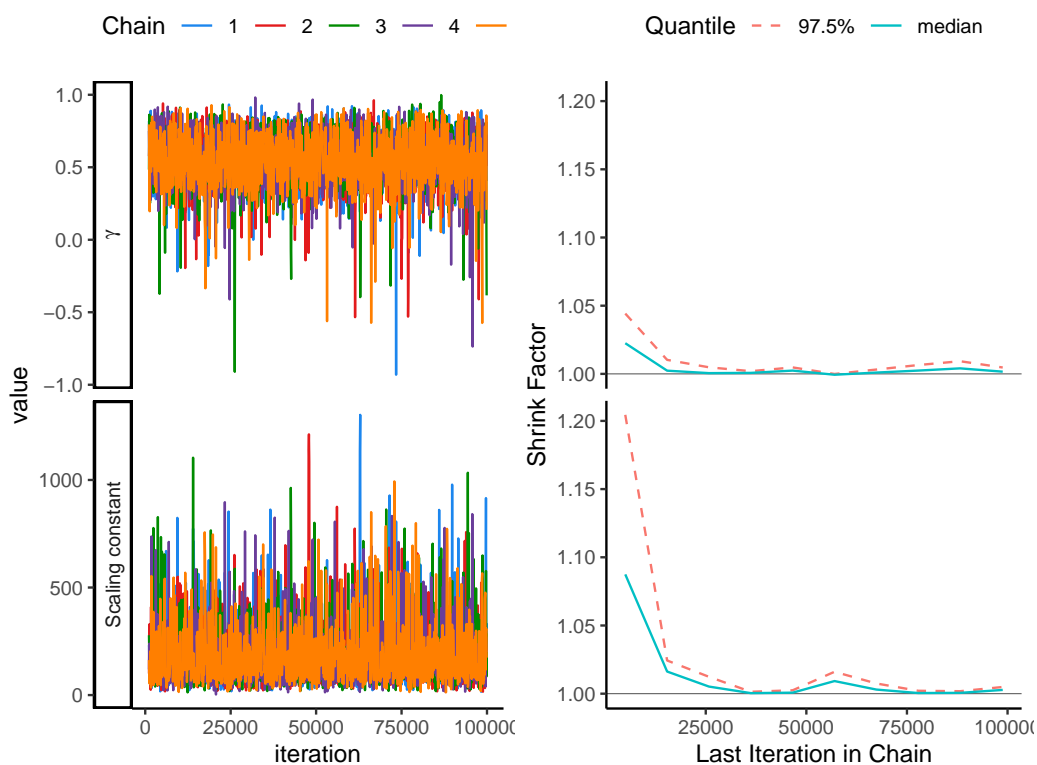


Figure S4: MCMC convergence diagnostics for the chaining model. On the left trace-plots, on the right changes along the chain of the Gelman and Rubin shrink factor

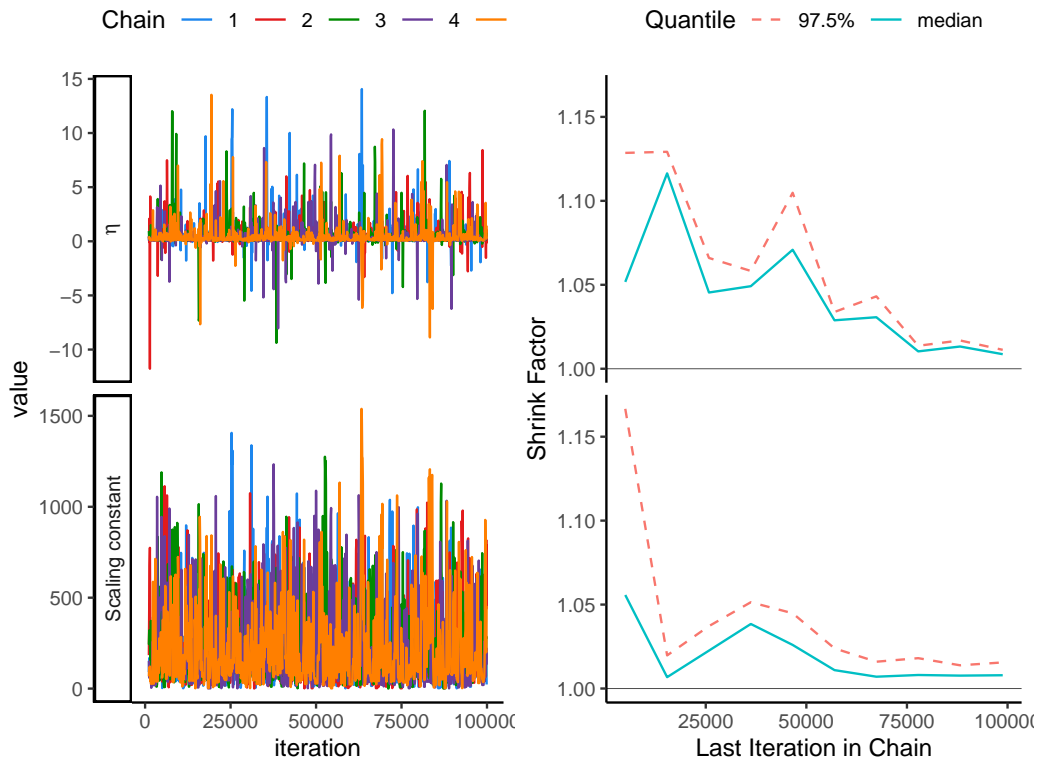


Figure S5: MCMC convergence diagnostics for the penalty model. On the left trace-plots, on the right changes along the chain of the Gelman and Rubin shrink factor

Table S1: Parameter values with which the model was run in the MCMC. σ refers to the amplitude of the perturbation kernel with the subscript indicating the associated parameter. New values were taken from a uniform distribution. α refers to the learning rate.

Parameter	Value
Learning rounds	10000
Reward value	1
α	0.05
σ_γ	0.3
σ_η	4
$\sigma_{Sca.Const.}$	300
Number of chains	5
Chain length	1^5

References

- Axén, Annkristin H., Olof Leimar, and Veronika Hoffman. 1996. “Signalling in a Mutualistic Interaction.” *Animal Behaviour* 52 (2): 321–33. <https://doi.org/10.1006/anbe.1996.0178>.
- Behrens, Timothy E. J., Laurence T. Hunt, Mark W. Woolrich, and Matthew

- 534 F. S. Rushworth. 2008. “Associative Learning of Social Value.” *Nature*
535 456 (7219): 245–49. <https://doi.org/10.1038/nature07538>.
- 536 Brandon, Susan E., Edgar H. Vogel, and Allan R. Wagner. 2002. “Compu-
537 tational Theories of Classical Conditioning.” In *A Neuroscientist’s Guide*
538 *to Classical Conditioning*, edited by John W. Moore, 232–310. New York,
539 NY: Springer. https://doi.org/10.1007/978-1-4419-8558-3_7.
- 540 Bshary, Redouan. 2001. “The Cleaner Fish Market.” In *Economics in Na-*
541 *ture: Social Dilemmas, Mate Choice and Biological Markets*, edited by
542 Jan A. R. A. M. Van Hooff, Peter Hammerstein, and Ronald Noë, 146–
543 72. Cambridge: Cambridge University Press. [https://doi.org/10.1017/](https://doi.org/10.1017/CBO9780511752421.010)
544 [CBO9780511752421.010](https://doi.org/10.1017/CBO9780511752421.010).
- 545 Bshary, Redouan, and Alexandra S Grutter. 2002. “Asymmetric Cheating
546 Opportunities and Partner Control in a Cleaner Fish Mutualism.” *Animal*
547 *Behaviour* 63 (3): 547–55. <https://doi.org/10.1006/anbe.2001.1937>.
- 548 Bshary, Redouan, and Rui F Oliveira. 2015. “Cooperation in Animals: To-
549 ward a Game Theory Within the Framework of Social Competence.” *Cur-*
550 *rent Opinion in Behavioral Sciences*, Social behavior, 3 (June): 31–37.
551 <https://doi.org/10.1016/j.cobeha.2015.01.008>.
- 552 Bshary, Redouan, and Daniel Schaffer. 2002. “Choosy Reef Fish Select
553 Cleaner Fish That Provide High-Quality Service.” *Animal Behaviour* 63
554 (3): 557–64. <https://doi.org/10.1006/anbe.2001.1923>.
- 555 Dally, Joanna M., Nathan J. Emery, and Nicola S. Clayton. 2006. “Food-
556 Caching Western Scrub-Jays Keep Track of Who Was Watching When.”

557 *Science (New York, N.Y.)* 312 (5780): 1662–65. [https://doi.org/10.1126/](https://doi.org/10.1126/science.1126539)
558 [science.1126539](https://doi.org/10.1126/science.1126539).

559 Demairé, Camille, Zegni Triki, Sandra A. Binning, Gaëtan Glauser, Do-
560 minique G. Roche, and Redouan Bshary. 2020. “Reduced Access to
561 Cleaner Fish Negatively Impacts the Physiological State of Two Resi-
562 dent Reef Fishes.” *Marine Biology* 167 (4): 48. [https://doi.org/10.1007/](https://doi.org/10.1007/s00227-020-3658-2)
563 [s00227-020-3658-2](https://doi.org/10.1007/s00227-020-3658-2).

564 Dubois, Thibault, Cristian Pasquaretta, Andrew B. Barron, Jacques Gau-
565 trais, and Mathieu Lihoreau. 2021. “A Model of Resource Partitioning
566 Between Foraging Bees Based on Learning.” *PLOS Computational Biol-*
567 *ogy* 17 (7): e1009260. <https://doi.org/10.1371/journal.pcbi.1009260>.

568 Enquist, Magnus, Johan Lind, and Stefano Ghirlanda. 2016. “The Power of
569 Associative Learning and the Ontogeny of Optimal Behaviour.” *Royal So-*
570 *ciety Open Science* 3 (11): 160734. <https://doi.org/10.1098/rsos.160734>.

571 Farashahi, Shiva, Katherine Rowe, Zohra Aslami, Daeyeol Lee, and Alireza
572 Soltani. 2017. “Feature-Based Learning Improves Adaptability Without
573 Compromising Precision.” *Nature Communications* 8 (1): 1768. <https://doi.org/10.1038/s41467-017-01874-w>.

574 <https://doi.org/10.1038/s41467-017-01874-w>.

575 Farashahi, Shiva, Jane Xu, Shih-Wei Wu, and Alireza Soltani. 2020. “Learn-
576 ing Arbitrary Stimulus-Reward Associations for Naturalistic Stimuli In-
577 volves Transition from Learning about Features to Learning about Ob-
578 jects.” *Cognition* 205 (December): 104425. [https://doi.org/10.1016/j.](https://doi.org/10.1016/j.cognition.2020.104425)
579 [cognition.2020.104425](https://doi.org/10.1016/j.cognition.2020.104425).

- 580 Galef, Bennett G. 1992. "The Question of Animal Culture." *Human Nature*
581 3 (2): 157–78. <https://doi.org/10.1007/BF02692251>.
- 582 Heyes, C. M. 1998. "Theory of Mind in Nonhuman Primates." *Behav-*
583 *ioral and Brain Sciences* 21 (1): 101–14. [https://doi.org/10.1017/](https://doi.org/10.1017/S0140525X98000703)
584 [S0140525X98000703](https://doi.org/10.1017/S0140525X98000703).
- 585 Heyes, Cecilia. 2012. "Simple Minds: A Qualified Defence of Associative
586 Learning." *Philosophical Transactions of the Royal Society B: Biological*
587 *Sciences* 367 (1603): 2695–2703. <https://doi.org/10.1098/rstb.2012.0217>.
- 588 Hirata, Satoshi, Kunio Watanabe, and Kawai Masao. 2001. "'Sweet-Potato
589 Washing' Revisited." In *Primate Origins of Human Cognition and Be-*
590 *havior*, edited by Tetsuro Matsuzawa, 487–508. Tokyo: Springer Japan.
591 https://doi.org/10.1007/978-4-431-09423-4_24.
- 592 Kawamura, Syunzo. 1959. "The Process of Sub-Culture Propagation Among
593 Japanese Macaques." *Primates* 2 (1): 43–60. [https://doi.org/10.1007/](https://doi.org/10.1007/BF01666110)
594 [BF01666110](https://doi.org/10.1007/BF01666110).
- 595 Kolodny, O., S. Edelman, and A. Lotem. 2014. "The Evolution of Con-
596 tinuous Learning of the Structure of the Environment." *Journal of The*
597 *Royal Society Interface* 11 (92): 20131091–91. [https://doi.org/10.1098/](https://doi.org/10.1098/rsif.2013.1091)
598 [rsif.2013.1091](https://doi.org/10.1098/rsif.2013.1091).
- 599 Macphail, Euan M. 1982. *Brain and Intelligence in Vertebrates*. Clarendon
600 Press.
- 601 Marler, Peter, and Susan Peters. 1989. "Species Differences in Auditory
602 Responsiveness in Early Vocal Learning." In *The Comparative Psychol-*

603 *ogy of Audition: Perceiving Complex Sounds*, 243–73. Hillsdale, NJ, US:
 604 Lawrence Erlbaum Associates, Inc.

605 McFadden, Daniel. 1974. “Conditional Logit Analysis of Qualitative Choice
 606 Behavior.” *Frontiers in Econometrics*, Frontiers in econometrics. - New
 607 York [u.a.] : Academic Press, ISBN 0-12-776150-0. - 1974, p. 105-142,.

608 Miller, George A. 1956. “The Magical Number Seven, Plus or Minus Two:
 609 Some Limits on Our Capacity for Processing Information.” *Psychological*
 610 *Review* 63 (2): 81.

611 Noë, Ronald, and Peter Hammerstein. 1995. “Biological Markets.” *Trends*
 612 *in Ecology & Evolution* 10 (8): 336–39. [https://doi.org/10.1016/S0169-](https://doi.org/10.1016/S0169-5347(00)89123-5)
 613 [5347\(00\)89123-5](https://doi.org/10.1016/S0169-5347(00)89123-5).

614 Prat, Yosef, Redouan Bshary, and Arnon Lotem. 2022. “Modelling How
 615 Cleaner Fish Approach an Ephemeral Reward Task Demonstrates a Role
 616 for Ecologically Tuned Chunking in the Evolution of Advanced Cognition.”
 617 *PLOS Biology* 20 (1): e3001519. [https://doi.org/10.1371/journal.pbio.](https://doi.org/10.1371/journal.pbio.3001519)
 618 [3001519](https://doi.org/10.1371/journal.pbio.3001519).

619 Quiñones, Andrés E., Olof Leimar, Arnon Lotem, and Redouan Bshary. 2019.
 620 “Reinforcement Learning Theory Reveals the Cognitive Requirements for
 621 Solving the Cleaner Fish Market Task.” *The American Naturalist*, De-
 622 cember, 000–000. <https://doi.org/10.1086/707519>.

623 R Core Team. 2021. “R: A Language and Environment for Statistical Com-
 624 puting.” Manual. Vienna, Austria. <https://www.R-project.org/>.

625 Rescorla, R. A., and A. R. A. Wagner. 1972. “A Theory of Pavlovian

626 Conditioning: Variations in the Effectiveness of Reinforcement and Non-
627 Reinforcement.” In *Classical Conditioning II: Current Research and The-*
628 *ory*, edited by Abraham H. Black and William Frederick Prokasy. New
629 York: Appleton-Century-Crofts.

630 Ros, Albert F. H., Jeanne Lusa, Meghann Meyer, Marta Soares, Rui F.
631 Oliveira, Michel Brossard, and Redouan Bshary. 2011. “Does Access to
632 the Bluestreak Cleaner Wrasse *Labroides Dimidiatus* Affect Indicators of
633 Stress and Health in Resident Reef Fishes in the Red Sea?” *Hormones and*
634 *Behavior* 59 (1): 151–58. <https://doi.org/10.1016/j.yhbeh.2010.11.006>.

635 Salwiczek, Lucie H., Laurent Prétôt, Lanila Demarta, Darby Proctor, Jen-
636 nifer Essler, Ana I. Pinto, Sharon Wismer, Tara Stoinski, Sarah F. Bros-
637 nan, and Redouan Bshary. 2012. “Adult Cleaner Wrasse Outperform
638 Capuchin Monkeys, Chimpanzees and Orang-Utans in a Complex Forag-
639 ing Task Derived from Cleaner – Client Reef Fish Cooperation.” *PLOS*
640 *ONE* 7 (11): e49068. <https://doi.org/10.1371/journal.pone.0049068>.

641 Schultz, Wolfram. 2015. “Neuronal Reward and Decision Signals: From
642 Theories to Data.” *Physiological Reviews* 95 (3): 853–951. [https://doi.](https://doi.org/10.1152/physrev.00023.2014)
643 [org/10.1152/physrev.00023.2014](https://doi.org/10.1152/physrev.00023.2014).

644 Shettleworth, Sara J. 2009. *Cognition, Evolution, and Behavior*. 2 edition.
645 Oxford ; New York: Oxford University Press.

646 Sih, Andrew, and Marco Del Giudice. 2012. “Linking Behavioural Syn-
647 dromes and Cognition: A Behavioural Ecology Perspective.” *Philosophical*
648 *Transactions of the Royal Society of London B: Biological Sciences*

367 (1603): 2762–72. <https://doi.org/10.1098/rstb.2012.0216>.

Staddon, J. E. R. 2016. *Adaptive Behavior and Learning*. 2 edition. Cambridge: Cambridge University Press.

Suddendorf, Thomas, and Michael C. Corballis. 2007. “The Evolution of Foresight: What Is Mental Time Travel, and Is It Unique to Humans?” *The Behavioral and Brain Sciences* 30 (3): 299–313; discussion 313–351. <https://doi.org/10.1017/S0140525X07001975>.

Sutherland, R. J., and J. W. Rudy. 1989. “Configural Association Theory: The Role of the Hippocampal Formation in Learning, Memory, and Amnesia.” *Psychobiology* 17 (2): 129–44. <https://doi.org/10.3758/BF03337828>.

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. Edited by Francis Bach. Second edition edition. Cambridge, MA: A Bradford Book.

Taborsky, Barbara, and Rui F. Oliveira. 2012. “Social Competence: An Evolutionary Approach.” *Trends in Ecology & Evolution* 27 (12): 679–88. <https://doi.org/10.1016/j.tree.2012.09.003>.

Triki, Zegni, Yasmin Emery, Magda C. Teles, Rui F. Oliveira, and Redouan Bshary. 2020. “Brain Morphology Predicts Social Intelligence in Wild Cleaner Fish.” *Nature Communications* 11 (1): 6423. <https://doi.org/10.1038/s41467-020-20130-2>.

Triki, Zegni, Alexandra S. Grutter, Redouan Bshary, and Albert F. H. Ros. 2016. “Effects of Short-Term Exposure to Ectoparasites on Fish Cortisol

672 and Hematocrit Levels.” *Marine Biology* 163 (9): 187. [https://doi.org/](https://doi.org/10.1007/s00227-016-2959-y)
673 10.1007/s00227-016-2959-y.

674 Triki, Zegni, Sharon Wismer, Elena Levorato, and Redouan Bshary. 2018.
675 “A Decrease in the Abundance and Strategic Sophistication of Cleaner
676 Fish After Environmental Perturbations.” *Global Change Biology* 24 (1):
677 481–89. <https://doi.org/10.1111/gcb.13943>.

678 Triki, Zegni, Sharon Wismer, Olivia Rey, Sandra Ann Binning, Elena Lev-
679 orato, and Redouan Bshary. 2019. “Biological Market Effects Predict
680 Cleaner Fish Strategic Sophistication.” *Behavioral Ecology* 30 (6): 1548–
681 57. <https://doi.org/10.1093/beheco/arz111>.

682 Varela, Susana A. M., Magda C. Teles, and Rui F. Oliveira. 2020. “The
683 Correlated Evolution of Social Competence and Social Cognition.” *Func-*
684 *tional Ecology* 34 (2): 332–43. <https://doi.org/10.1111/1365-2435.13416>.

685 Waldie, Peter A., Simon P. Blomberg, Karen L. Cheney, Anne W. Goldizen,
686 and Alexandra S. Grutter. 2011. “Long-Term Effects of the Cleaner Fish
687 *Labroides Dimidiatus* on Coral Reef Fish Communities.” *PLOS ONE* 6
688 (6): e21201. <https://doi.org/10.1371/journal.pone.0021201>.

689 Wismer, Sharon, Ana I. Pinto, Zegni Triki, Alexandra S. Grutter, Dominique
690 G. Roche, and Redouan Bshary. 2019. “Cue-Based Decision Rules of
691 Cleaner Fish in a Biological Market Task.” *Animal Behaviour*, October.
692 <https://doi.org/10.1016/j.anbehav.2019.09.013>.

693 Wismer, Sharon, Ana I. Pinto, Alex L. Vail, Alexandra S. Grutter, and
694 Redouan Bshary. 2014. “Variation in Cleaner Wrasse Cooperation and

695 Cognition: Influence of the Developmental Environment?" *Ethology* 120
696 (6): 519–31. <https://doi.org/10.1111/eth.12223>.