

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, ex-

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

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

From an evolutionary perspective, mechanisms are likely selected because of how they allow individuals to respond to environmental variation. For example, biological market theory predicts that the exchange rate of goods and/or services traded between cooperative partners adjusts to the law of supply 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
95 multiple cleaning stations while others access the only cleaning station in
96 their territory, it is crucial to categorize clients as either “visitors” or “resi-
97 dents”, respectively. During cleaning interactions, a cleaner fish often faces
98 a choice between a visitor and a resident client seeking its cleaning services
99 simultaneously. Visitors have the option to switch to another cleaner fish if
100 being made to wait, while residents must wait for inspection. Indeed, visitors
101 have been observed to use their partner choice option in that way (Bshary
102 and Schäffer 2002), which may explain why cleaners give visitors service pri-
103 ority in a field study in the Red Sea (Bshary 2001). Furthermore, in a lab
104 based paradigm, design to mimic the resident-visitor choice (ephemeral re-
105 ward task), cleaners learned to prefer the cue associated with the epheral

106 food source (visitor) and hence accessed both food sources, obtaining double
107 the amount of food (Bshary and Grutter 2002). However, further exploration
108 revealed that not all cleaner fish manage to develop a preference for
109 the ephemeral option in the lab (Triki et al. 2018). Over the last decade,
110 over a hundred wild-caught cleaner fish have been tested in the exact same
111 paradigm of the ephemeral reward task (Salwiczek et al. 2012; Wismer et
112 al. 2014; Triki et al. 2018, 2019, 2020). These fish often come from different
113 reef locations. Investigation of the local eco-sociological conditions revealed
114 that cleaner and client fish population densities have a substantial impact
115 on cleaner fish performance in the task. Cleaner fish from reef sites with
116 relatively low densities were more likely to fail at solving the task (Triki et
117 al. 2019, 2018; Wismer et al. 2014). This intra-specific variation is unlikely
118 due to local genetic adaptation, because cleaner fish are open water spawners
119 and the environmental 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 conditioning)
122 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 residents,
128 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

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

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

172 **Methods**

173 **The model**

174 The model consists of a set of individual-based simulations where individ-
175 uals, representing cleaner fish, face a series of choices between two options,
176 which simulate the natural conditions of the cleaning market. Individuals ex-
177 perience a series of discrete time points in which they face different ‘states’,
178 defined by the number and category of client fish (visitor or residents) invit-
179 ing for cleaning services. There are six possible states: zero clients, one res-
180 ident, 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 difference 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
223 of two parameters that could also work together. First, γ measures how
224 much individuals include future rewards in their decision updates. If $\gamma = 0$,
225 individuals only use the immediate reward obtained from a cleaning interac-
226 tion. As γ increases, individuals include more of the reward obtained from
227 the subsequent choices. That amounts to estimating and using for decision
228 making the future expected rewards of an action (chaining). Second, η mea-
229 sures how much individuals include in their reward the fleeing behavior of
230 visitors as a negative component (penalty). Both of these parameters allow
231 individuals 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 visi-
 239 tors leaving if made to wait, and the performance of wild-caught cleaner fish
 240 in the ephemeral reward test. In total, we have twelve site/year data sets for
 241 fish censuses and corresponding performance in the lab test. Thus, some sites
 242 were sampled more than once. To estimate the population density of cleaner
 243 fish and their clients at a given site in a given year, Triki et al. (2019) used a
 244 series of ten transects of $30m$ each. Observers swam along the transect lines
 245 placed on the reef and first counted the visible large-bodied adult fish (species
 246 with total length $TL \geq 10cm$) including cleaner fish on a width of $5m$, and
 247 then on the return individuals of small-bodied fish species ($TL < 10cm$) on
 248 a width of $1m$ (see Triki et al. (2018) for further details on fish censuses
 249 data collection). Total length estimates were done by the observer. We then
 250 scaled the counts of cleaner fish, small-bodied, and large-bodied clients fish
 251 densities per $100m^2$.

252 The field observation data consisted of video recordings/encodings of the
 253 cleaner-client cleaning interactions. There were videos from eight cleaners
 254 per site/year of a duration of $30min$ each. Triki et al. extracted information
 255 from every event wherein a visitor client was made to wait in favor of another
 256 client (visitor or resident), and noted whether or not the visitor left or queued
 257 for the cleaning service (Triki et al. 2019, 2020).

258 The cognitive performance data was from a total of 120 cleaners (10 individ-
 259 uals per 12 site/year) tested in the ephemeral reward task (Triki et al. 2019,
 260 2020). Authors housed all captured cleaners individually in glass aquaria
 261 ($62cm \times 27cm \times 37cm$) and provided them with PVC pipes ($10cm \times 1cm$) as
 262 shelters. The task consisted of exposing the cleaner fish to substitute models
 263 of client fish in the form of two *Plexiglas* plates offering the same amount
 264 of food (one item of mashed prawn). The two plates differed in colour and
 265 pattern (horizontal green stripes or vertical pink stripes) but had equal size
 266 ($10cm \times 7cm$). Importantly, the two plates played different roles as either a
 267 visitor (ephemeral food source) or resident (permanent food source). That is,
 268 if a cleaner fish inspected the resident plate first, the experimenter withdrew
 269 the visitor plate out of the aquarium as a consequence. Choosing first the
 270 visitor plate, however, granted access to both plates. The equal size of the
 271 plates forced cleaner fish to decide based solely on the association between
 272 the behaviour and the color/pattern cue (Wismer et al. 2019). Triki et al.
 273 (2019, 2020) tested the fish for a maximum of 200 trials with 20 trials a day,
 274 10 trials in the morning and 10 trials in the afternoon. They randomized and
 275 counterbalanced the plates' spatial location (i.e. left or right) between trials.
 276 Similarly, they counterbalanced the plates' decoration (colour and pattern)
 277 and the plates' role (visitor or resident) between the tested fish. In the
 278 original studies, once a fish reached a learning criterion, that is, performing
 279 significantly above chance level in a binomial test ($p - value \leq 0.05$), they
 280 passed to a reversal version of the task where the roles of the visitor/resident

281 Plexiglas plates were swapped. The reversal phased stopped when the fish
 282 performed significantly above chance, or the fish completed the 200 trial to-
 283 gether with the initial; see Triki et al.(2019, 2020). Here, we are interested in
 284 explaining the total frequency of visitor choices using the model, rather than
 285 just the achievement of the criterion. Total frequency of visitor choices nat-
 286 urally comes out of the model, and allows us to use all the variation among
 287 cleaners, instead of reducing that to a binomial variable. Thus, we used
 288 instead a subset of these data to estimate the final cleaner fish preferences
 289 for the visitor plate, even if they do not reach the learning criterion within
 290 200 trials. To do so, we first extracted the trial-by-trial outcomes from the
 291 last two sessions (20 trials) of those who never reached the learning criterion
 292 for visitor plate ($N = 45$ cleaner fish). For those who reached the learning
 293 criterion at some point during the test and passed to a reversal phase, we
 294 extracted the trial-by-trial outcomes from the last session (10 trials) before
 295 passing to reversal and the last 10 trials they were exposed to in the test
 296 ($N = 75$ cleaner fish).

297 We chose a combination of initial and reversal to quantify preference for the
 298 visitor client. However, it could be argued that using only the initial phase
 299 gives a better estimation of the cleaner fish preference for the visitor. In
 300 the supplementary material (Fig. S1) we show how using initial and reversal
 301 (a), and only initial (b) maps to the previously used criteria. The initial and
 302 reversal match better the criteria chosen in previous analysis of the ephemeral
 303 reward task experimental set-up (Triki et al. 2019, 2020).

304 Statistical analysis

305 The aim of the analyses is to fit the key model parameters γ and η , to the
306 empirical data from Triki et al. (2019, 2020) to test whether each or a
307 combination of these effects is a better explanation for the pattern seen in
308 the data. We used the ecological variables: cleaners, visitor clients, resident
309 clients abundances and visitor clients leaving probability,
310 as input to the models. As the response variable, we used the frequency
311 with which cleaners chose the visitor option in the ephemeral reward task.
312 Finally, we used the probability with which agents in the model simulations
313 choose the visitor in resident-visitor options as the prediction for the response
314 variable. We kept all other parameter values used for the model simulations
315 constant, see Table S1.

316 To capture with the model the relationship between the ecological variables
317 and cleaner fish preferences for visitors, we needed to scale the absolute
318 population densities of cleaner fish from the empirical data to a measure of
319 relative abundances that captures client visitation patterns. This is because,
320 in the model, relative abundances of clients define not only the probability
321 of residents and visitors but also how often the cleaning station is empty
322 (e.g. there are no clients to be cleaned). The frequency with which clients
323 visit the station is another variable influencing the station occupancy, which
324 in nature may vary among different client species depending on their ectopar-
325 asite loads. We do not have field estimates for species-specific parasite loads,

326 especially not as a function of the site. In order to control for these aspects,
 327 we computed a measure of relative cleaner fish abundance for each reef site
 328 relative to absolute abundances and multiplied it by a scaling constant that
 329 changes the range of the variable. Specifically, cleaner fish relative abun-
 330 dance is equal to $\frac{\epsilon C_{abs}}{\epsilon C_{abs} + R_{abs} + V_{abs}}$. Where ϵ is the scaling constant; C_{abs} is
 331 the cleaner fish absolute density; R_{abs} is the residents absolute density and
 332 V_{abs} is the visitor client absolute density. This scaling constant is meant to
 333 capture variation in the market conditions driven partly by cleaner fish abun-
 334 dance. We fitted the value of the scaling constant (ϵ) as part of the statistical
 335 inference. As for the visitor and resident abundances, we computed a relative
 336 measure with respect to the total client abundance, and weighted that by the
 337 re-scaled relative absence of cleaners ($1 - \text{relative cleaner abundance}$). Thus,
 338 all three measures of relative abundance sum up to 1, and can be used in
 339 the model as a proxy for the probability of having different options in the
 340 cleaning station. Note that we introduced the scaling constant to control for
 341 variation that is not captured by the model; its parameter distribution does
 342 not offer biological insights.

343 Once we calculated the relative abundances, we obtained predictions from
 344 the model for each one of the locations and ran the Markov Chain. We
 345 started the chain with random values for the parameters of interest, then
 346 ran the computational model once for every reef site using as input the eco-
 347 logical explanatory variables. The model outputs the probability of choosing
 348 the visitor for each location p_i , where i is the index for the 12 locations.

349 Assuming a binomial distribution, the probability that each of the cleaner
 350 fish 20 choices in the ephemeral reward task was generated by the model is
 351 given 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
 352 fish j in reef site i chose the visitor over the resident; and $n = 20$ (due to our
 353 choice of using 20 choices per cleaner fish). By taking the natural logarithm
 354 and summing over all the cleaner fish and reef sites, we obtained the log-
 355 likelihood of the data given the model and parameters. We then proposed a
 356 new set of parameters drawn from a uniform distribution centered around the
 357 old parameter set. The amplitude of the uniform distribution used for each
 358 parameter can be found in Table S1. Subsequently, we ran the model and
 359 calculated the likelihood with the new parameter set. We then used the ratio
 360 of the two likelihoods to choose which parameter set to keep. New parameter
 361 sets with a higher likelihood than the old set replaced old ones, and those
 362 with a lower likelihood replaced current ones with a probability equal to the
 363 log-likelihood ratio. Given that we only used the likelihoods in the decision,
 364 we used an uninformative prior. Once we decided whether the new param-
 365 eter set would replace the old one, we ran the model again to sample the
 366 likelihood distribution of the parameter set. We then started the cycle again
 367 by proposing a new set of parameters and repeated the process for $1e^5$ steps.
 368 We ran 5 independent chains, discarded the first 1000 samples of each chain
 369 as burn in, and after that, we kept 1 in every 100 samples to avoid autocorre-
 370 lation. The collection of parameters sets kept in all chains approximates the
 371 posterior distribution. We coded the model as well as the fitting algorithm in

372 `c++`, the diagnostics and visualization in R (R Core Team 2021). All codes
373 are accessible at <https://zenodo.org/badge/latestdoi/440585701>

374 To compare the fit of the three alternative models, we used the distribution of
375 pseudo R^2 proposed by Mcfadden (McFadden 1974). Mcfadden’s pseudo R^2
376 is a standard measure of fit for logistic regression. In that context, *pseudo*—
377 R^2 uses the log-likelihood of the data given the model, relative to the log-
378 likelihood of the data given a model without covariates, as a measure of fit.
379 Our model is not a logistic regression, therefore we measured the *pseudo*— R^2
380 in relation to the log-likelihood of a model with parameters γ and η set to
381 zero. This, in practice, amounts to a model that has a neutral preference
382 between the two options. We computed the pseudo R^2 for all the samples of
383 the posterior from the MCMC. We used these distributions of pseudo R^2 ’s
384 as a measure of fit.

385 Results

386 Estimation of parameter values for the three models (full model, chaining
387 and penalty), support chaining as the only mechanism cleaners use to ac-
388 count for the future effects of their actions; and thus to solve the ephemeral
389 reward task. In the estimation of the parameter values of the full model,
390 which includes both chaining and penalty, the bulk of the marginal posterior
391 distribution of η which controls the strength of penalty is around 0 (Fig.
392 1). As for γ , controlling chaining, the 95% confidence intervals also includes

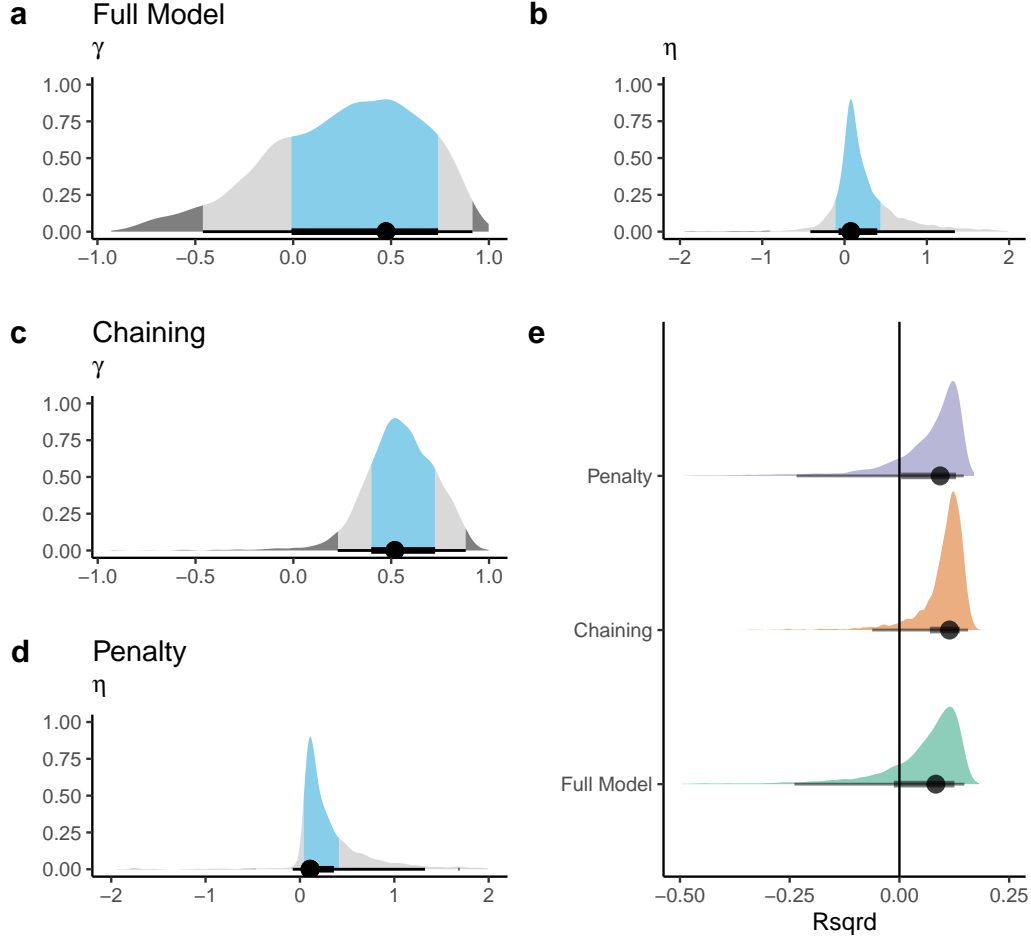


Figure 1: Posterior distributions for parameter values γ and η for the three models (Full model, chaining only and penalty only). We show the kernel density estimates, below the mode (black dot) and the 66 (light blue shade) and 95% (grey shade) highest posterior density interval for the two parameters. 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.

393 zero, but the mode of the posterior is around 0.5 (Fig. 1, a). In the chaining
 394 model, where η is set to zero, the distribution of γ shifted to higher values,
 395 zero is no longer part of the 95% credible interval of the parameter (Fig.
 396 1, c). In contrast, when we look at the model with only penalty, the pos-
 397 terior distribution of η is still centered around zero (Fig. 1, d). Thus, the
 398 analysis of the estimates of individual parameter values in the three models
 399 only supports a strong effect of chaining. Furthermore, the comparison of
 400 the models' fit favors the chaining model. In panel e of figure 1 we show the
 401 distribution of $pseudo - R^2$ calculated using samples from the posterior dis-
 402 tributions shown before. Note, $pseudo - R^2$ can have negative values, which
 403 is when the log-likelihood of the model is lower than that of a model that
 404 triggers neutral preferences. Even though the peak of the three $pseudo - R^2$
 405 distributions were not very different, the model with only chaining produced
 406 a distribution of $pseudo - R^2$ where more values were positive (to the right
 407 of the black line in Fig. 1 e). This shows that accounting for variation in the
 408 parameter estimates the model with chaining gives a better fit to the data,
 409 despite having one parameter less than the full model. We have not shown
 410 here the marginal posterior distributions of the scaling constant, given that
 411 they do not bring biological insight. Their visualizations can be found in the
 412 supplementary material (Fig. S2), as well as the diagnostics of the MCMCs
 413 (Figs. S3,S4,S5).

414 Low cleaner fish relative abundance yields different predictions from a model
 415 with only chaining and with penalty. The model with only chaining predicts

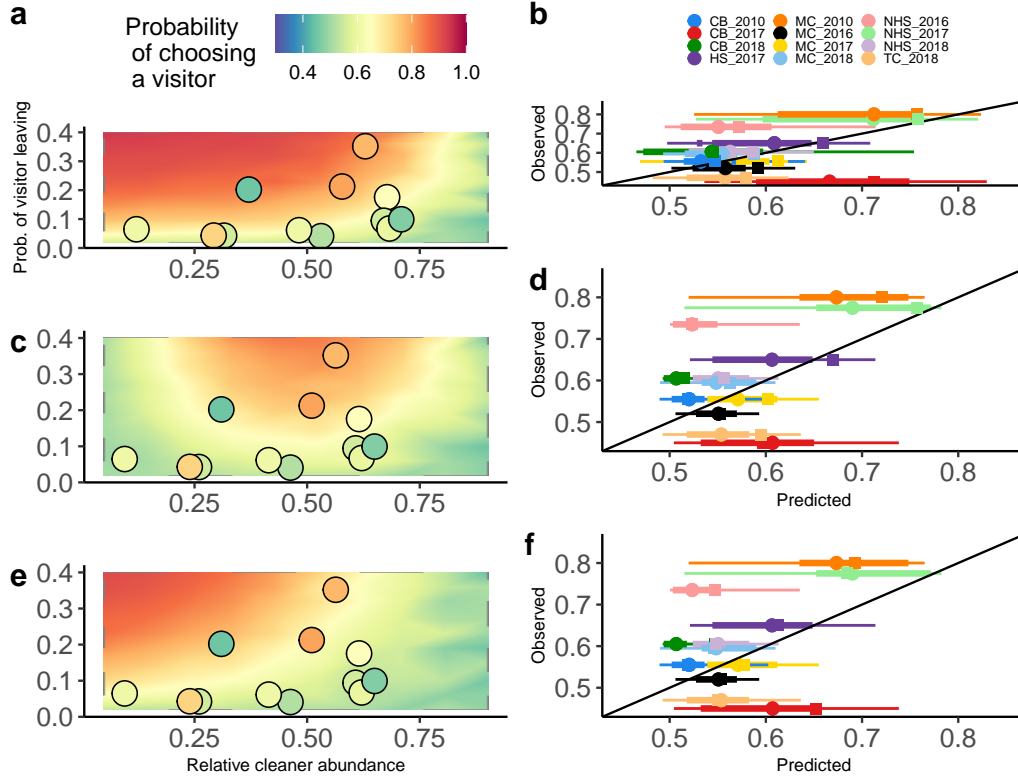


Figure 2: Observed and predicted probability of choosing a visitor. Left-hand side panel: colour contour shows the prediction from the learning model using the mode of the posterior distributions of parameters recovered by the statistical analysis. Dots show the frequency of visitor choices for the 12 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.

416 a low preference for the visitor when cleaner fish relative abundance is low
417 (Fig. 2 c). In contrast, the models including penalty (with and without
418 chaining) predict high preference for the visitor option (Fig. 2 a and e).
419 For intermediate and high cleaner relative abundances the predictions are
420 similar for all three models. Intermediate relative abundance trigger a high
421 preference for the visitor; while low abundances trigger low preference for
422 the visitor (Fig. 2a,b,e). Note, however, we calculated preferences shown in
423 figure 2 left panels by using only the model of the posterior distributions, and
424 by holding constant the balance between resident and visitors' abundances.
425 Panels on the right, show how close predictions are from the observed data,
426 allowing the balance between client types to vary and using a set of samples
427 from the posterior distribution.

428 Visitor leaving probability has an overall positive effect on the probability of
429 choosing the visitor clients on all three models. All three models predict an
430 increase preference for visitors as the visitor probability increases (Fig. 2).

431 Discussion

432 In this study, our main aim was to unravel which of two potential cognitive
433 mechanisms, chaining of events, penalty, or their combination, best explains
434 wild-caught cleaner fish performance in the ephemeral reward task, while
435 accounting for their ecological conditions. To evaluate the merits of each
436 of these two mechanisms separately and combined, we considered cleaner

437 fish performance in the lab test to have its origin from the rule these fish
438 applied in their natural environment. That is, individuals that solved the
439 task already had a preference for visitor clients and generalized this rule to
440 the lab conditions once being familiar with the task.

441 While all three models captured well the positive relationship between visitor
442 leaving behavior and cleaner fish performance in the market task (Triki et
443 al. 2019), only the chaining mechanism predicted that cleaner fish perfor-
444 mance in the task should be low in habitats with low cleaner-to-client ratios,
445 regardless of the visitor leaving probability. In contrast, models including
446 negative reward predicted the highest performance in the ephemeral reward
447 task when relative cleaner fish abundance is low, particularly together with
448 a high probability of visitor leaving (Fig. 2). Low relative cleaner fish abun-
449 dances mean the market has an excess of demand for cleaning services. In
450 the models, this translates to a cleaning station that is frequently full. Thus,
451 when a visitor leaves, it is likely that the cleaner fish will have access to
452 another client in the next step. Therefore, there will not be much difference
453 in future reward between choosing a visitor and a resident, and cleaners will
454 not develop a preference for the visitor in these conditions. On the other
455 hand, the effect of negative reward on cleaner fish preference is the opposite,
456 as in a busy cleaning station, to that of chaining. Cleaner fish will get more
457 often the resident-visitor state and will develop a preference for the visitor
458 faster. At high cleaner fish abundances, the resident-visitor state becomes so
459 rare that neither mechanism is very efficient at generating a preference for

460 visitors. When facing the resident-visitor choice, it is still best to choose the
461 visitor; however, the learning machinery will not be able to develop this pref-
462 erence efficiently. Overall, the models suggest that chaining is the cognitive
463 mechanism that allows cleaner fish to adaptively adjust to their biological
464 market ecological conditions.

465 Previous research showed that cleaner fish living at high population densi-
466 ties and giving service priority to the visitor plate in the ephemeral reward
467 task, as well as cleaner fish living at low densities but denying service pri-
468 ority to the visitor plate possess larger forebrains; a key teleost brain region
469 associated with behavioural flexibility and social intelligence. Those failing
470 to show optimized decision-rules given their local ecological conditions had
471 relatively smaller forebrains (Triki et al. 2020). Triki et al. refer to the for-
472 mer as socially competent cleaner fish, while the second group as socially
473 incompetent cleaner fish. Social competence is the ability to optimise social
474 behavior to the available social information (Taborsky and Oliveira 2012;
475 Bshary and Oliveira 2015; Varela, Teles, and Oliveira 2020). Our analyses
476 yielded no evidence that the difference in social competence with respect
477 to the local ecological conditions and associated brain morphology, found
478 by Triki et al. (2020), is due to the mechanism used to incorporate future
479 consequences. It is conceivable that high performing individuals from low
480 population densities reef sites use negative reinforcement instead of chain-
481 ing, but in that case, negative reinforcement should have explained at least
482 part of the data. Configurational learning or chunking (Sutherland and Rudy

1989; Miller 1956), the second component necessary to solve the ephemeral reward task (Quiñones et al. 2019), was not varied in the models we analysed here. However, while chunking tendencies should vary to allow individuals to adapt to their local conditions (Prat, Bshary, and Lotem 2022; Kolodny, Edelman, and Lotem 2014), systematic differences in individual chunking tendencies would not explain how socially competent decisions vary as a function of relative abundance. Therefore, it remains currently unclear what cell-demanding mechanisms may cause variation in social competence that translates into site-specific variation in performance in the ephemeral reward 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

506 the development of preferences for morphological or behavioural features that
507 are statistically associated with visitors or residents. For example, visitors
508 are on average larger than residents in body size (Bshary 2001), and contrary
509 to residents, they are less likely to chase a cleaner fish that fails to cooperate
510 and instead cheats its client by taking a bite of mucus (Bshary and Grutter
511 2002). Given these associations, chaining might produce the decision-rule
512 “choose the larger client and/or the less aggressive client”, which is not a
513 useful rule in the standard ephemeral reward task.

514 In conclusion, our study shows that variation in cognitive performance as a
515 function of the local ecological conditions may set the stage for the use of
516 mechanistic modelling to identify the cognitive processes underlying learning
517 in animals. The combination overcomes the limitations of the general philos-
518 ophy in animal cognition to apply the logic of Lloyd Morgan’s canon (Occam’s
519 razor). Cognitive experiments with the aim of excluding basic reinforcement
520 learning as a potential explanation (operant and/or classical conditioning)
521 of performance often employ one trial experiments requiring animals to solve
522 the task on the first possible occasion. For example, any theory of mind
523 task needs to be solved in the first trial in order to exclude fast condition-
524 ing (C. M. Heyes 1998). Similarly, subjects need to solve a social learning
525 task on the first trial to accept imitation as a mechanism over stimulus/local
526 enhancement. Such strict conditions are virtually never met. For example,
527 potato washing by Japanese macaques, an iconic example of social learning,
528 took several years to spread within the group (Kawamura 1959), meaning

529 that any learner had been repeatedly exposed to demonstrations before ac-
530 quisition. Importantly, Galef (1992) refuted imitation as a mechanism not
531 simply because of the repeated exposure but because a (rather qualitative)
532 analysis of the spread of potato washing across individuals did not follow
533 the prediction based on imitation learning (see also Hirata, Watanabe, and
534 Masao (2001)). In our case, the number of trials it took cleaners to learn
535 the solution to the ephemeral rewards task would never allow excluding an
536 important role of penalty based on the data alone. However, fitting model
537 predictions to our comprehensive empirical data set revealed that a more
538 complex mechanism, estimation of future reward, fits the data better.

Supplementary material

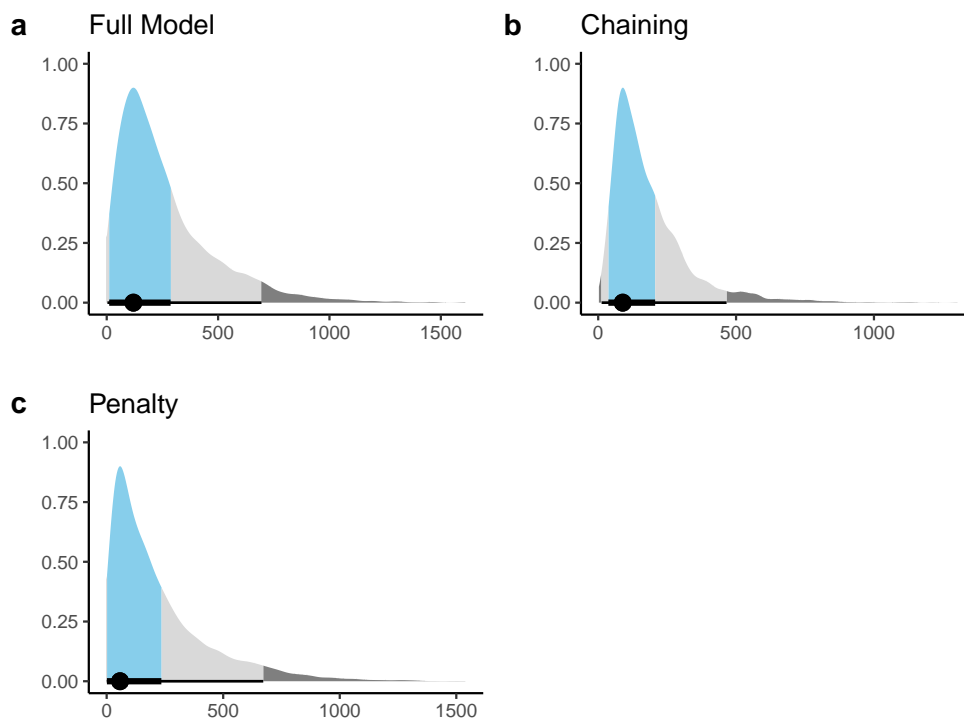


Figure S2: Posterior distributions for scaling constant for the three models (Full model, chaining only and penalty only). We show the kernel density estimates, below the mode (black dot) and the 65% (light blue shade) and 95% (grey shade) highest posterior density interval. On the top, panel a shows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

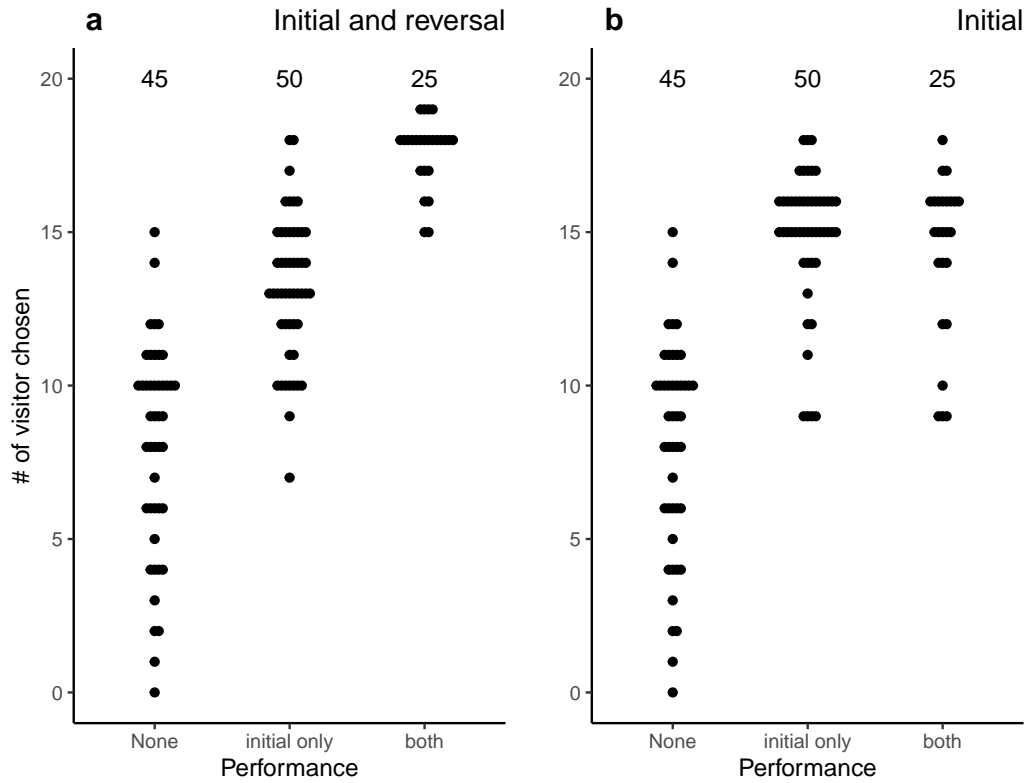


Figure S1: Relation between the response variable used in this study and the criteria used in previous studies to assess performance in the ephemeral reward task. In the x axis, we classified the performance of cleaner fish according to whether they developed a preference for the visitor in the initial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

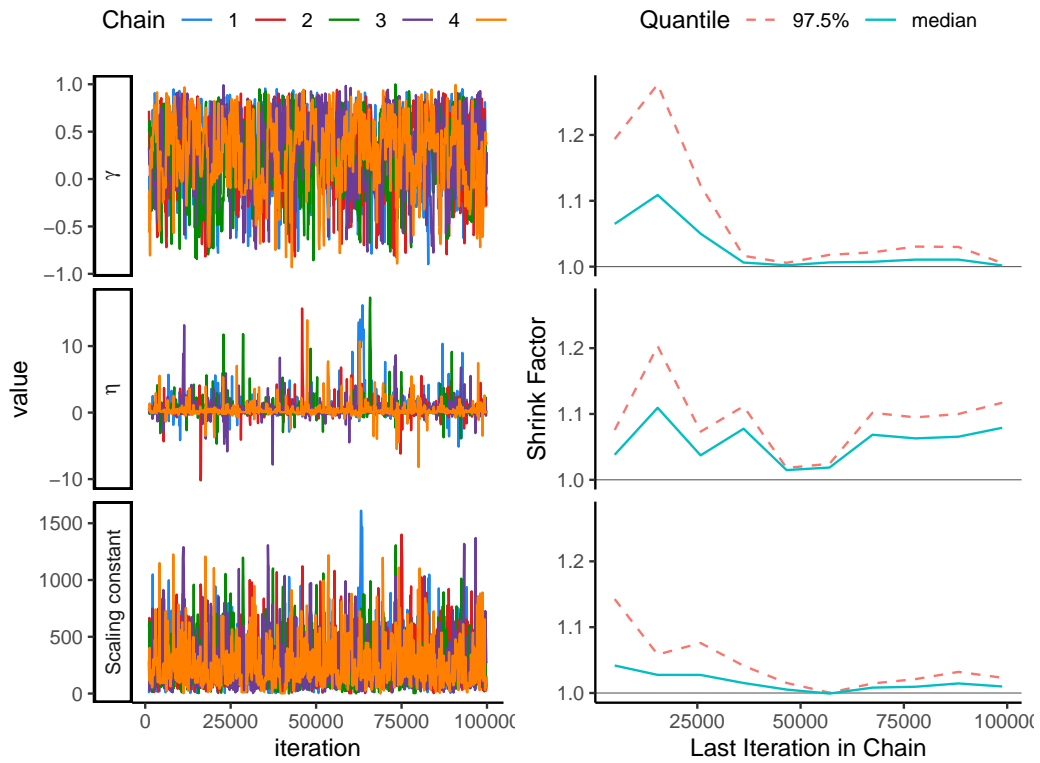


Figure S3: MCMC convergence diagnostics for the full model. On the left trace-plots, on the right changes along the chain of the Gelman and Rubin shrink factor.

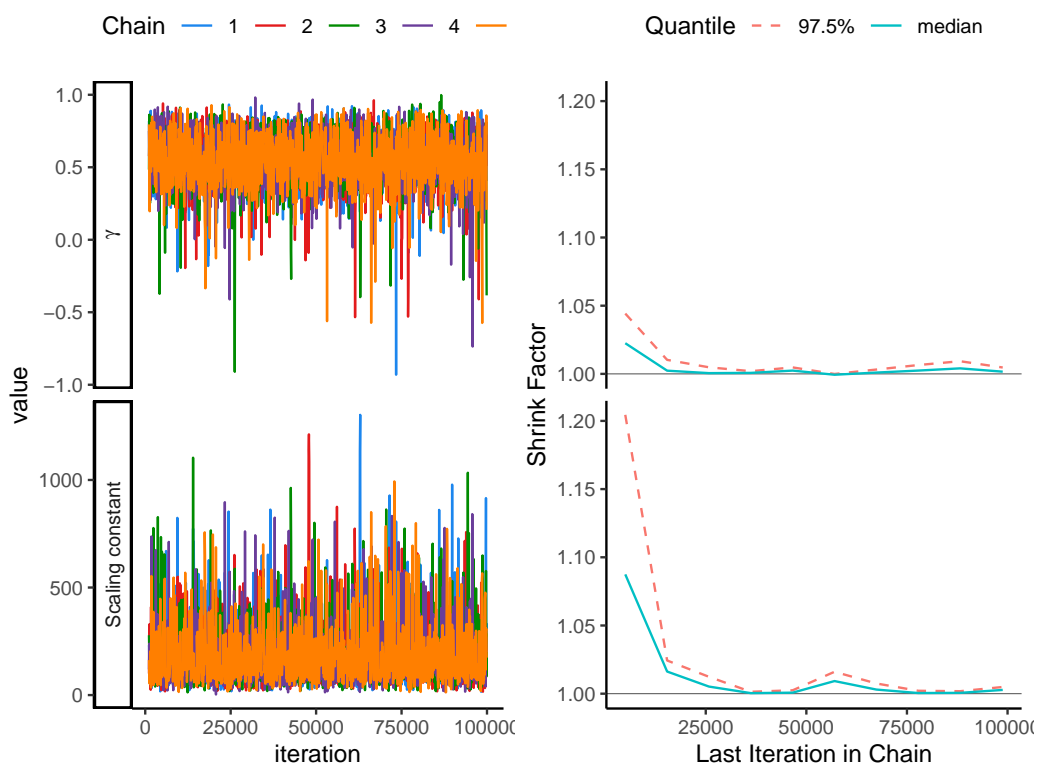


Figure S4: MCMC convergence diagnostics for the chaining model. On the left trace-plots, on the right changes along the chain of the Gelman and Rubin shrink factor

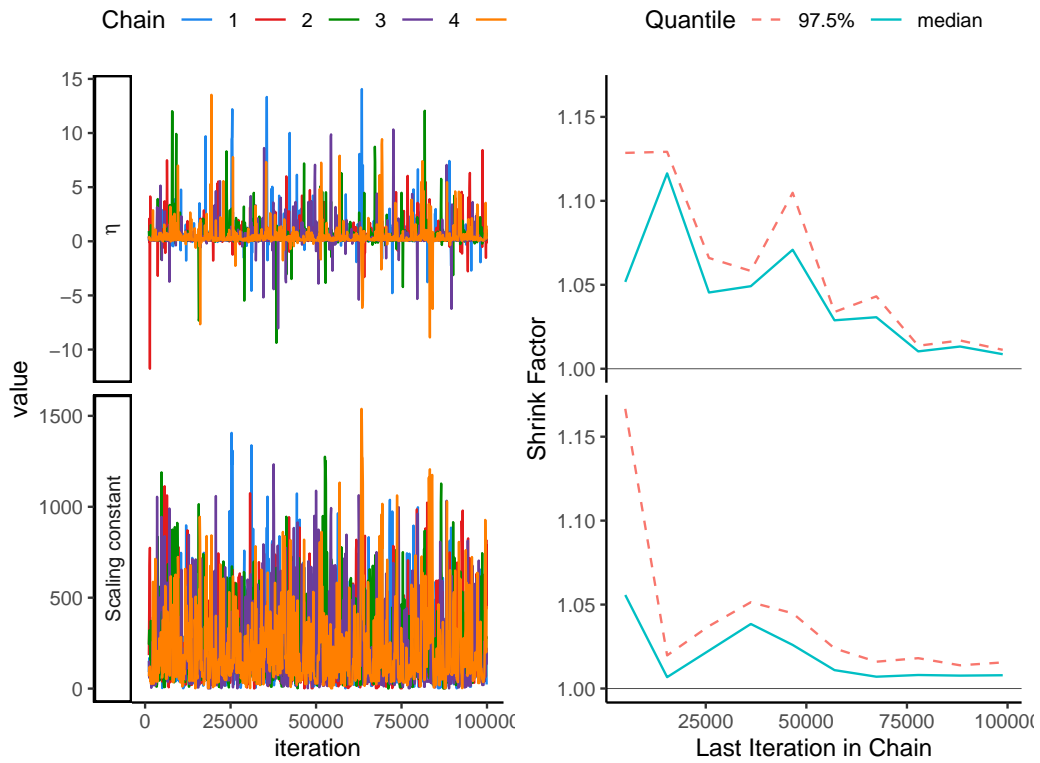


Figure S5: MCMC convergence diagnostics for the penalty model. On the left trace-plots, on the right changes along the chain of the Gelman and Rubin shrink factor

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

- 545 F. S. Rushworth. 2008. “Associative Learning of Social Value.” *Nature*
546 456 (7219): 245–49. <https://doi.org/10.1038/nature07538>.
- 547 Brandon, Susan E., Edgar H. Vogel, and Allan R. Wagner. 2002. “Compu-
548 tational Theories of Classical Conditioning.” In *A Neuroscientist’s Guide*
549 *to Classical Conditioning*, edited by John W. Moore, 232–310. New York,
550 NY: Springer. https://doi.org/10.1007/978-1-4419-8558-3_7.
- 551 Bshary, Redouan. 2001. “The Cleaner Fish Market.” In *Economics in Na-*
552 *ture: Social Dilemmas, Mate Choice and Biological Markets*, edited by
553 Jan A. R. A. M. Van Hooff, Peter Hammerstein, and Ronald Noë, 146–
554 72. Cambridge: Cambridge University Press. [https://doi.org/10.1017/](https://doi.org/10.1017/CBO9780511752421.010)
555 [CBO9780511752421.010](https://doi.org/10.1017/CBO9780511752421.010).
- 556 Bshary, Redouan, and Alexandra S Grutter. 2002. “Asymmetric Cheating
557 Opportunities and Partner Control in a Cleaner Fish Mutualism.” *Animal*
558 *Behaviour* 63 (3): 547–55. <https://doi.org/10.1006/anbe.2001.1937>.
- 559 Bshary, Redouan, and Rui F Oliveira. 2015. “Cooperation in Animals:
560 Toward a Game Theory Within the Framework of Social Competence.”
561 *Current Opinion in Behavioral Sciences*, Social behavior, 3 (June): 31–37.
562 <https://doi.org/10.1016/j.cobeha.2015.01.008>.
- 563 Bshary, Redouan, and Daniel Schaffer. 2002. “Choosy Reef Fish Select
564 Cleaner Fish That Provide High-Quality Service.” *Animal Behaviour* 63
565 (3): 557–64. <https://doi.org/10.1006/anbe.2001.1923>.
- 566 Dally, Joanna M., Nathan J. Emery, and Nicola S. Clayton. 2006. “Food-
567 Caching Western Scrub-Jays Keep Track of Who Was Watching When.”

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

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

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

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

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

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

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

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

614 *of Audition: Perceiving Complex Sounds*, 243–73. Hillsdale, NJ, US:
615 Lawrence Erlbaum Associates, Inc.

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

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

622 Noë, Ronald, and Peter Hammerstein. 1995. “Biological Markets.” *Trends*
623 *in Ecology & Evolution* 10 (8): 336–39. [https://doi.org/10.1016/S0169-](https://doi.org/10.1016/S0169-5347(00)89123-5)
624 5347(00)89123-5.

625 Prat, Yosef, Redouan Bshary, and Arnon Lotem. 2022. “Modelling How
626 Cleaner Fish Approach an Ephemeral Reward Task Demonstrates a Role
627 for Ecologically Tuned Chunking in the Evolution of Advanced Cogni-
628 tion.” *PLOS Biology* 20 (1): e3001519. [https://doi.org/10.1371/journal.](https://doi.org/10.1371/journal.pbio.3001519)
629 pbio.3001519.

630 Quiñones, Andrés E., Olof Leimar, Arnon Lotem, and Redouan Bshary.
631 2019. “Reinforcement Learning Theory Reveals the Cognitive Require-
632 ments for Solving the Cleaner Fish Market Task.” *The American Natu-*
633 *ralist*, December, 000–000. <https://doi.org/10.1086/707519>.

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

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

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

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

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

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

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

657 Sih, Andrew, and Marco Del Giudice. 2012. “Linking Behavioural Syn-
658 dromes and Cognition: A Behavioural Ecology Perspective.” *Philosophical
659 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

683 and Hematocrit Levels.” *Marine Biology* 163 (9): 187. [https://doi.org/](https://doi.org/10.1007/s00227-016-2959-y)
684 10.1007/s00227-016-2959-y.

685 Triki, Zegni, Sharon Wismer, Elena Levorato, and Redouan Bshary. 2018.
686 “A Decrease in the Abundance and Strategic Sophistication of Cleaner
687 Fish After Environmental Perturbations.” *Global Change Biology* 24 (1):
688 481–89. <https://doi.org/10.1111/gcb.13943>.

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

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

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

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

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

706 Cognition: Influence of the Developmental Environment?" *Ethology* 120
707 (6): 519–31. <https://doi.org/10.1111/eth.12223>.