A role for ecologically-tuned chunking in the evolution of advanced cognition demonstrated by modelling the cleaner fish market problem — Source link 🔗

Yosef Prat, Redouan Bshary, Lotem A

Institutions: Tel Aviv University

Related papers:

- Domain learning naming game for color categorization.

- Chunks are not "Content-Free": Hierarchical Representations Preserve Perceptual Detail within Chunks

- TaxoNERD: deep neural models for the recognition of taxonomic entities in the ecological and evolutionary literature

- An investigation of the mechanism of information reduction

- Advanced community question answering by leveraging external knowledge and multi-task learning

1

A role for ecologically-tuned chunking in the evolution of advanced cognition demonstrated by modelling the cleaner fish market problem

4

5   Yosef Prat[1*], Redouan Bshary[1], Arnon Lotem[2]

6   [1] Institute of Biology, University of Neuchâtel, Emile-Argand 11, 2000, Neuchâtel, Switzerland.

7   [2] School of Zoology, Tel Aviv University, Tel Aviv, Israel.

8   * Corresponding author: yosefprat@gmail.com

9

10   **Abstract:** What makes cognition 'advanced' is an open and not precisely defined question. One perspective involves increasing the complexity of associative learning, from conditioning to learning sequences of events ('chaining') to representing various cue combinations as 'chunks'. Here we develop a weighted-graph model to study the conditions for the evolution of chunking ability, based on the ecology of the cleaner fish *Labroides dimidiatus*. Cleaners must learn to serve visitor clients before resident clients, because a visitor leaves if not attended while a resident waits for service. This challenge has been captured in various versions of the ephemeral-reward task, which has been proven difficult for a range of cognitively capable species. We show that chaining is the minimal requirement for solving the laboratory task, that involves repeated simultaneous exposure to an ephemeral and permanent food source. Adding ephemeral-ephemeral and permanent-permanent combinations, as cleaners face in the wild, requires individuals to have chunking abilities to solve the task. Importantly, chunking parameters need to be calibrated to ecological conditions in order to produce adaptive decisions. Thus, it is the fine tuning of this ability which may be the major target of selection during the evolution of advanced associative learning.

24

25

26

Introduction

In an effort to understand the evolution of cognition, a wide range of studies has been focused on identifying cognitive abilities in animals that appear "advanced", and exploring the ecological conditions that could possibly favour their evolution (e.g., [1–6]). Yet, mapping sophisticated cognitive abilities along phylogenetic trees and their relation to social or ecological conditions (e.g., [7,8]) does not explain how such abilities evolved to become "advanced" through incremental modifications of their mechanistic building blocks. Earlier views of cognitive evolution were based on some postulated, loosely defined genetic adaptations, such as language instinct [9,10], mind reading abilities [11], or mirror neurons [12,13], but those are increasingly replaced by approaches relying on explicit associative learning principles that can gradually form complex representations of statistically learned information [14–22]. In line with these recent views, in order to understand the critical steps in cognitive evolution, one should identify specific modifications that can elaborate simple learning processes and make them better in some way.

A relatively simple and well-understood example is the extension of simple conditioning through second-order conditioning in a process known as chaining [23,24]. In this process, a stimulus associated with a primary reinforcer (such as a sound associated with receiving food), becomes a reinforcer by itself, and then a stimulus reinforced by the new reinforcer may become a reinforcer, and so on, allowing to represent sequences of statistical dependencies. Such sequences could, in turn, facilitate navigation [25,26] or even social learning [27].

Further elaborations of associative learning that may allow to construct complex representations and to support advanced forms of statistical learning and decision-making are less well-understood. It has become clear, however, that a critical ability required for such cognitive advances is the ability to represent two or more data units as a *chunk* or a *configuration* that has a meaning that is different from (or independent of) the meaning of its components. This ability has appeared in the literature under different names, such as *configurational learning* [28,29], *chunking* [19,30,31], or *segmentation* [32], all of which are quite similar, and involve the learning of configurations, patterns, and hierarchical structures in time and space [33].

In its simple form, known as configurational learning, this ability allows to learn, for example, that the elements A and B are associated with positive reward while their configuration AB is not rewarded and should therefore be avoided (a task known as negative patterning [34]). Configurational learning of this type is contrasted with *elemental learning*, which is based on the behaviour expected from simple associative learning [35,36]. Research on configurational learning has been focused mainly on identifying the brain regions supporting this ability (e.g., [29,37–39]), giving relatively little attention to the cognitive processes generating configural representations (but see [29]). More attempts to consider these possible processes has been made in the context of chunking or segmentation (e.g.

2

62   [14,32,40]), but only recently, theoretical work has started to address the question of how chunking

63   mechanisms evolve under different ecological conditions, and what is their role in cognitive evolution

64   [19,41,42].

65   A unique model system that may provide a remarkable opportunity to study the evolution of chunking

66   is that of the bluestreak cleaner wrasse (*Labroides dimidiatus*) which feeds on ectoparasites removed

67   from 'client' fish [43]. Field observations and laboratory experiments have shown that at least some of

68   these cleaner fish are capable of solving a problem known as the market problem (or the ephemeral

69   reward task) [44–47]. The market problem entails that if approached by two clients, cleaners must learn

70   to serve a visitor client before a resident client, because the latter waits for service while the former

71   leaves if not attended (see details in [44,46,48,49]). In the lab, clients of different types were replaced

72   with plates of different colours, each offering one food item and made to act like either a visitor or a

73   resident [45,50]. Interestingly, individuals captured in different habitats demonstrated different learning

74   abilities of the market problem in the lab, and adult cleaners seem to learn better than juveniles

75   [46,48,51,52]. Such intraspecific variation in cognitive abilities suggests some role for the ecological

76   and the developmental circumstances in the fish life history.

77   The lab market task may first appear as a two-choice experiment, testing whether animals can learn to

78   choose the option that yields the largest total amount of food. Nevertheless, while preferring a larger

79   amount in a simple two-choice task seems almost trivial for most animals [53], the market version, in

80   which a double amount is a product of a sequence of two actions (i.e. choosing a visitor and then

81   approaching the resident) has been proven difficult for a range of species [52,54–56] (but see [57–59]).

82   Follow-up studies on pigeons and rats (reviewed in [60]) showed that letting the subject make a first

83   decision but delaying the consequences, i.e. delaying the access to the rewarding stimuli, strongly

84   improves performance [60,61]. One interpretation of these results is that the delay helped animals to

85   connect their initial choice to both consequences; the first, and then the second reward, both of which

86   occurred within a short time span after the relatively long delay.

87   While delaying the consequences of the initial choice may be helpful under some conditions, recent

88   theoretical work suggests that under natural conditions, basic associative learning is insufficient for

89   solving the market task, which instead warrants some form of chunking ability [62]. The reason for that

90   is that the commonly used laboratory task presents a relatively simple version of the problem compared

91   to the natural situation. It only presents visitor-resident pairs, for which choosing the visitor first, always

92   entails double rewards and choosing the resident first always entails a single reward. In nature, on the

93   other hand, cleaners face also resident-resident as well as visitor-visitor pairs, and most often only a

94   single client approaches. As a result, choosing visitor first may not always entail double reward (e.g.,

95   in visitor-visitor pairs the second visitor is likely to leave) and choosing resident first may not always

96   result in a single reward (e.g., in resident-resident pairs the second resident is likely to stay). Indeed,

97   the theoretical analysis carried out by Quiñones et al. [62] showed that for solving the natural market

3

98   problem it is necessary to have distinct representations of all different types of client combinations
99   (visitor (*v*) + resident (*r*), *r+r*, *v+v*, *r*, and *v*), which means the ability to represent chunks. Yet, the
100   analyses did not explain how such representations are created, and to what extent ecology causes
101   variation in the cleaners' ability to create such representations.

102   Following Quiñones et al.'s demonstration that chunking is necessary for solving the natural market
103   problem, here we use the cleaner fish example as a means to study the evolution of an explicit chunking
104   mechanism and the conditions that favour its success. Thus, we investigate how the very same problem
105   – choosing between two options where one yields the double amount of food – set into an increasingly
106   complex ecology selects for the evolution of increasingly advanced associative learning abilities. Our
107   model is based on a weighted directed graph of nodes and edges, which initially form a simple
108   associative learning model, and can then be modified to become an extended credit (chaining-like)
109   model, or a chunking model. This approach allows to compare between clearly defined learning
110   mechanisms and to pinpoint the modifications responsible for a presumed evolutionary step that
111   improves cognitive ability. We analyse the three learning models' performance in three tasks: the basic
112   quantitative choice task, the laboratory market task and the market task embedded in a sequence of
113   varying configurations ('natural market task'). For the latter, we explored to what extent different
114   densities and frequencies of client types select for different tendencies to form chunks (a critical
115   parameter in the model), and how such different tendencies may affect the cleaners' ability to solve the
116   market problem.

117

The core model

Internal representation

120   Our core model consists of a weighted directed graph $G = (N, E)$, with nodes N, edges E, and
121   additionally edge-weights W, node-weights U, and node values F (Fig. 1A). The basic model includes
122   three internal nodes representing three external cues (perceived states): $N = \{V, R, X\}$, where: V –
123   serving (feeding on) a visitor-client, R – serving a resident-client, and X – absence of clients (empty
124   arena). These are the three cues required to represent the market problem and are therefore available to
125   the cleaner fish in our simulations (Fig. 2). Edge-weights are updated according to the sequential
126   appearance of the cues, i.e., whenever n_j appears after n_i the weight of the edge $n_i \rightarrow n_j$, i.e. $W(n_i, n_j)$,
127   increases (by one unit, in our simulations). Thus, edge-weights represent the associative strength
128   between nodes experienced one after the other. Node weights and values are attached to the cleaner's
129   decisions (see decision-making below) according to their occurrence and association of their outcome
130   with food, i.e., whenever node n_i is chosen, the weight $U(n_i)$ increases (by one unit, in our simulations)
131   and the value $F(n_i)$ increases by the amount of food reward provided (which unless otherwise specified,

132 is assumed to be one unit per client if served successfully, and zero otherwise). The value of a node

133 could be regarded as the strength of its association with food, which can also be represented as the

134 weight of the edge between the node and a reinforcer food-node (the weights of green arrows in Fig.

135 1A). The weighted directed graph constitutes the cleaner's internal representation of the market

136 environment. The cleaner's decisions regarding which clients to serve depend only upon this

137 representation (Fig. 1A).

138 Initially, the cues are considered unknown to the fish and their corresponding values, weights, and the

139 weights of their connecting edges are set to zero. Most learning models use prior values for cues, which

140 are commonly set to zero (often implicitly). Here, we model such a prior by imposing a threshold on

141 the weight of a node before any increase in its value F can occur. Specifically, $F(n_k)$ is initialized to

142 zero, and would not change as long as $U(n_k) < Q$, i.e., at the first Q occurrences of n_k. We set $Q = 10$

143 throughout all simulations, which implies that the value of a node will increase above zero only from

144 the 11$^{\text{th}}$ serving of a client.

145
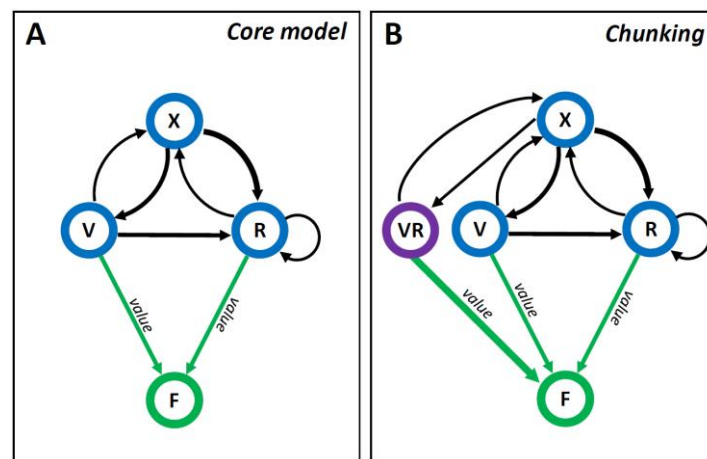
Figure 1. Model design - internal representation. **A**) The *core model* contains a network of three elements (blue circles) representing perceived states: *V* – serving a visitor-client, *R* – serving a resident-client, *X* – absence of clients. The value of each node is represented by the weight of its association (width of green arrows) with the reinforcer (food reward; green circle). Edge weights (width of black arrows) represent the strength of the associations between sequential states. This is also the internal representation of the *extended credit model*. **B**) An example of a possible representation in the *chunking model*: a new element (*VR*; purple circle) represents the configuration (the chunk) of '*V and then R*'.

146

Decision making

When a cleaner fish is presented with two clients, it must choose which one to serve first. If both clients are of the same type (i.e., v (visitor) and v, or r (resident) and r) the cleaner chooses one with equal probabilities. However, when two contrasting types are present (i.e., v and r) the decision is made according to the values associated with serving each type, $F(R)$ and $F(V)$. A soft-max function is employed (see [62]) over the *normalized values* $f(n_i)$ and $f(n_j)$ such that the probability of choosing n_i is:

$$\pi_i = \frac{1}{1 + e^{-(f(n_i) - f(n_j))}} \qquad \text{(Eq. 1)}$$

Where $f(n_k) = \frac{F(n_k)}{U(n_k)}$ is the average payoff associated with the node n_k. Note that the numerator, $F(n_k)$, is the sum of all obtained reward items associated with the cue n_k (i.e., the accumulated number of food items obtained after the cleaner has chosen n_k), and the denominator is a count of all occurrences of n_k (i.e., the number of times the cleaner has chosen n_k, regardless of whether this choice had been fulfilled).

The probability of choosing n_j is $\pi_j = 1 - \pi_i$.

In the market problems presented here, both client types provide the same immediate reward. Thus, it is quite intuitive that learning only first order associations cannot provide any discrimination between them, and consequently, would fail in developing a preference for visitors (which is the essence of solving the market problem). Indeed, as we shall see in the results section, the core model was never successful in solving the market problem (either in its simple laboratory version or more complex natural setting). Yet, it serves as a null model and as a stepping-stone for the more advanced learning models.

A linear operator model

To compare our core model with a similar known benchmark we used the linear-operator learning model [63], which is a basic and widely used learning model [64] that does not involve chaining or chunking. The learner updates the value *f(i)* of cue *i* at time *t* such that $f(i)_t = (1 - \alpha)f(i)_{t-1} + \alpha\varphi(i)_t$, where $\varphi(i)_t$ is the reward attached to cue *i* at time *t* and α is a learning-rate parameter. To choose between clients based on their updated values we used the same soft-max decision-making rule applied by the core model (see above).

177

The extended credit model

179   A straightforward approach to consider higher order dependencies is to enable association of cues with

180   their 'future' rewards. We call this model the 'extended credit' model. The network representation of

181   the extended credit model is the same as that of the core model (Fig. 1A) but in this model the learner

182   associates an obtained reward with the current cue as well as with the previous one. Specifically, while

183   encountering a sequence (n_i, n_j), if n_i is rewarding then $F(n_i)$ increases, and if n_j is rewarding then both

184      $F(n_i)$ and $F(n_j)$ increase (i.e. the credit assignment of the reward is extended also to the previous cue).

185   Hence, if both cues are similarly rewarding the first one will be associated with double the food by the

186   end of the sequence, as it was also associated with a delayed reward. Theoretically, credit assignment

187   could be extended in more than one step backward and the credit could also change (e.g., decrease) with

188   time (similarly to 'chaining' [65]). Note that although the model extends the credit to a previous cue, it

189   does not represent, in the credit extension, the **identity** of the consecutive cue which donated the extra

190   reward. Thus, the extended credit model cannot learn to distinguish between different sequences

191   (sequential combinations or configurations) of cues (e.g., $V{\to}R$, $V{\to}X$, $R{\to}R$, etc.). The decision-

192   making process of the extended credit model is the same as in the core model (see above).

193

The chunking model

195   Another way of identifying high order dependencies is via configurational learning, or chunking, as

196   mentioned in the Introduction section. To model how acquired experience leads individuals to create

197   chunks, we employ a chunking procedure in our model in which sequences occurring more often than

198   expected, according to the distribution of their elements are 'chunked' into a new element (Fig. 1B).

199   Specifically, a sequence (n_i, n_j) would become a new element 'n_in_j' of the internally-represented

200   network (i.e., a new node in the graph G) whenever:

$$W(n_i, n_j) > M \cdot P(n_i) \cdot P(n_j) + C_p \cdot \hat{\sigma} \qquad \text{(Eq. 2)}$$

201   where $W(n_i, n_j)$ is the number of observed occurrences of the sequence $n_i \to n_j$, M is the total number

202   of observed cues (or pair sequences), $P(n_k)$ is the observed frequency of the element n_k, and $\hat{\sigma}$ is the

203   standard deviation of a binomial distribution, with the probability of an event $n_i \to n_j$ being

204      $P(n_i)P(n_j)$:

$$\hat{\sigma} = \sqrt{M \cdot P(n_i)P(n_j) \cdot \left(1 - P(n_i)P(n_j)\right)} \qquad \text{(Eq. 3)}$$

205      $C_p \geq 0$ is a chunking avoidance parameter. This parameter is important, as it governs the behaviour of

206   our model, or in other words, the conditions under which a chunk will be created. Note that when $C_p{=}0$,

207   any slight above chance co-occurrence of n_i and n_j would result in chunking. This is probably too much

208   chunking because it can easily happen in nature for almost any two elements as a result of stochastic

209   deviations from the frequency expected by chance. Using a C_p that is greater than zero implies that a

210   chunk will be created only when the co-occurrence is higher than expected by a certain threshold.

211   Additionally, chunks would not be created as long as $W(n_i, n_j) < Q$, i.e., during the first Q occurrences

212   of the sequence $n_i \rightarrow n_j$. This rule enforces a minimal sample size before statistical inference could be

213   done (for chunking).

214   In this model (see Fig. 1B), whenever a chunk is created it is treated as a new node and is being

215   associated with food whenever chosen by the cleaner alongside food reward (but only after its first Q

216   occurrences, as required for other elements). For instance, if the sequence $V \rightarrow R$ is chunked into a new

217   element 'VR', further choices of the sequence $V \rightarrow R$ will increase the association of the element 'VR'

218   with the reward by two units (as this is the observed reward during the processing of the sequence). On

219   the other hand, if the sequence $R \rightarrow V$ is chunked into a new element 'RV' (which could happen in the

220   *natural market problem*; see simulated environments below), further choices of the sequence $R \rightarrow V$

221   will usually increase the association of the element 'RV' with the reward by one unit only (since the

222   visitor leaves if not served first).

223   The decision-making process of the chunking model is the same as in the core model (see above) but

224   here, more choices may become available. For example, after the chunk 'VR' is created, a cleaner faced

225   with a visitor and a resident client simultaneously can choose to serve the resident (R), or to perceive

226   them as the chunk 'VR' and to execute the sequence $V \rightarrow R$ (i.e. approach the visitor and then the

227   resident). On the other hand, if the chunk 'RV' was also created, an additional option exists, which is

228   the choice of executing the sequence $R \rightarrow V$. Importantly, in this case, soon after approaching the

229   resident, the visitor would leave the arena so the outcome of choosing and attempting to execute the

230   sequence $R \rightarrow V$ may end up with serving only R (depending on the simulated environment; see below)

231   and being reward by only one unit (see above). We assume that if a chunk has already been created the

232   cleaner never chooses the first element alone if presented with both elements (i.e., if 'VR' is already

233   represented in the network, and 'RV' is not, the cleaner should only choose between 'R' and 'VR' when

234   presented with both client types simultaneously).

235

Simulated environments

237   Our simulations provide the cleaner fish with alternating clients awaiting its service (Fig. 2). The

238   simulated arena includes two available spots, where each can be occupied by a visitor client, a resident

239   client, or remain empty (simultaneous encounters with more than two clients are relatively rare in nature

240   and are typically not addressed; see [44,45]). Each simulation consists of a sequence of discrete trials.

8

241   On each trial, the arena is filled using a random sample according to the simulation specific setup (i.e.,

242   the probabilities of encountering each client type, as will be explained below). When the cleaner is

243   presented with an empty arena (none of the two spots is occupied) it perceives the corresponding cue X

244   (see Fig. 1A) and waits for the next trial (the next occupancy of the two spots). When the cleaner is

245   presented with only a single client it immediately serves it, perceives the corresponding cue (i.e., V for

246   choosing to serve a visitor, or R for choosing to serve a resident), experience the associated reward of

247   serving it, and waits for the next trial. When the cleaner is presented with two clients, it chooses one

248   according to the decision rule of the model (see above) if the clients are of different types (a visitor and

249   a resident), or at random (with equal probabilities) if they are of the same type. The cleaner then serves

250   the chosen one, and perceives the corresponding cue (i.e., V or R) and its associated reward. If the

251   second client (not chosen) is a visitor it leaves and the cleaner waits for the next trial, but if the second

252   client is a resident, the cleaner serves it as well, perceives the corresponding cue (R) and its associated

253   reward, and waits for the next trial. Recall that whenever the cleaner chooses to serve a client and

254   perceives its associated food reward, it adds one unit to the value F of the corresponding cue (e.g., to

255   *F(V), F(R), F(VR)*, etc.).

256   We have simulated three different environments: *i*) A laboratory environment with a 'basic two-choice

257   task', where a cleaner has to choose between two clients offering a reward of 1 and 2 units respectively,

258   and no further approach to clients is allowed after this initial choice within a trial (only in this

259   simulation, both client types are ephemeral). This two-choice task is expected to be solved by all types

260   of leaners (i.e., preferring the client offering the double amount of food), thus serving as a ground-level

261   associative learning test. *ii*) A laboratory environment with a 'laboratory market problem', where the

262   cleaner faces a resident and a visitor client together (Fig. 2A), in each feeding trial, and after it finishes

263   feeding it faces a single trial of empty arena (Fig. 2F; i.e., perceives the cue X). *iii*) A natural setting,

264   henceforth termed 'the natural market problem', in which the cleaner may face all possible

265   combinations (Fig. 2A-F): a visitor and a resident, two residents, two visitors, a single client (resident

266   or visitor), and no clients. In addition, in the natural setting the cleaner does not necessarily have to wait

267   between trials. This environment simulates more faithfully the situation in the wild, where each of the

268   two spots is filled using an independent random sample, with a probability P_V for a visitor, a probability

269   P_R for a resident, and a probability P_0 for an empty spot ($P_V + P_R + P_0 = 1$). When examining *the*

270   *natural market problem,* we consider the distribution of the different client types and their combinations

271   as resulting from two ecological parameters: the (relative) visitor frequency, $\frac{P_V}{P_V+P_R}$ (the fraction of

272   visitors out of all clients), and the overall client density, $1 - P_0$ (ranging from zero – when there are no

273   clients and the cleaner always faces an empty arena, to one – the arena is always full).

274

275

9

276
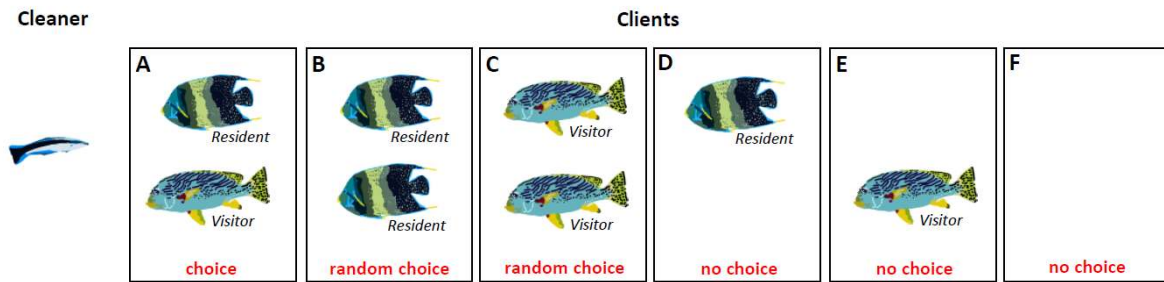
Figure 2. Model simulations. The cleaner in our simulations may encounter different combinations of client pairs awaiting its service: **A**) the cleaner must choose between two clients of different types according to the model's decision process, **B and C**) The cleaner chooses with equal probabilities between two clients of the same type, **D and E**) the cleaner serves the only available client, and **F**) the cleaner waits for clients to visit its cleaning station.

277

278

Results

280 We examined how the four learning models fare in the three simulated tasks: the basic two-choice task,
281 the laboratory market problem, and the natural market problem.

282

The basic two-choice task

284 All learning models solved successfully the basic two-choice task, as expected, exhibiting clear
285 preference for the client offering double amount of reward, and showing virtually no differences in
286 speed and accuracy of learning (Fig. 3A). In this task, there are no sequences of rewarding cues (as only
287 the chosen client is consumed and the other leaves) thus the advanced models are practically reduced
288 to the core model. Thus, the extended credit and the chunking model confer no extra benefit when facing
289 a basic two-choice test between options that differ in the amount of reward.

290

The laboratory market problem

292 Facing the laboratory market problem, the core model and the equivalent linear-operator learner did not
293 develop a preference towards any of the given clients and thus failed to solve the problem (Fig. 3B,
294 orange and blue lines, respectively). In contrast, both the extended-credit model and the chunking model
295 were capable of solving the *laboratory market problem*, i.e., to develop a strong preference towards the
296 visitor client (Fig. 3B, yellow and purple lines, respectively). The inability of the core and linear-
297 operator models to solve the laboratory market problem is reflected in their indifferent *normalized*

298   *value*s of R, $f(R)$, and V, $f(V)$, both of which approach 1 (Figure 3C). This result is expected since

299   the value of each cue is updated independently of any past and future cue or reward, and both cues

300   (client types) provide the exact same immediate reward. On the other hand, in the extended credit model

301   that solves the problem successfully, the *normalized value* of R, $f(R)$, approaches 1, while the

302   *normalized value* of V, $f(V)$, approaches 2 (Fig. 3D, top panel). This was made possible because serving

303   a resident always provides a single food item in this setup (as the visitor leaves) while the credit of

304   choosing a visitor is extended to the resident that waits to be served (thus crediting V with two food

305   items). The success of the chunking model is based on a different process: it creates the chunk *VR* early

306   and $f(VR)$ quickly approaches 2 as the complete chunk provides two food items (Fig. 3D, bottom

307   panel), pushing the preference towards a visitor client (the model choses the sequence *VR*, i.e., *V and*

308   *then R*).

309
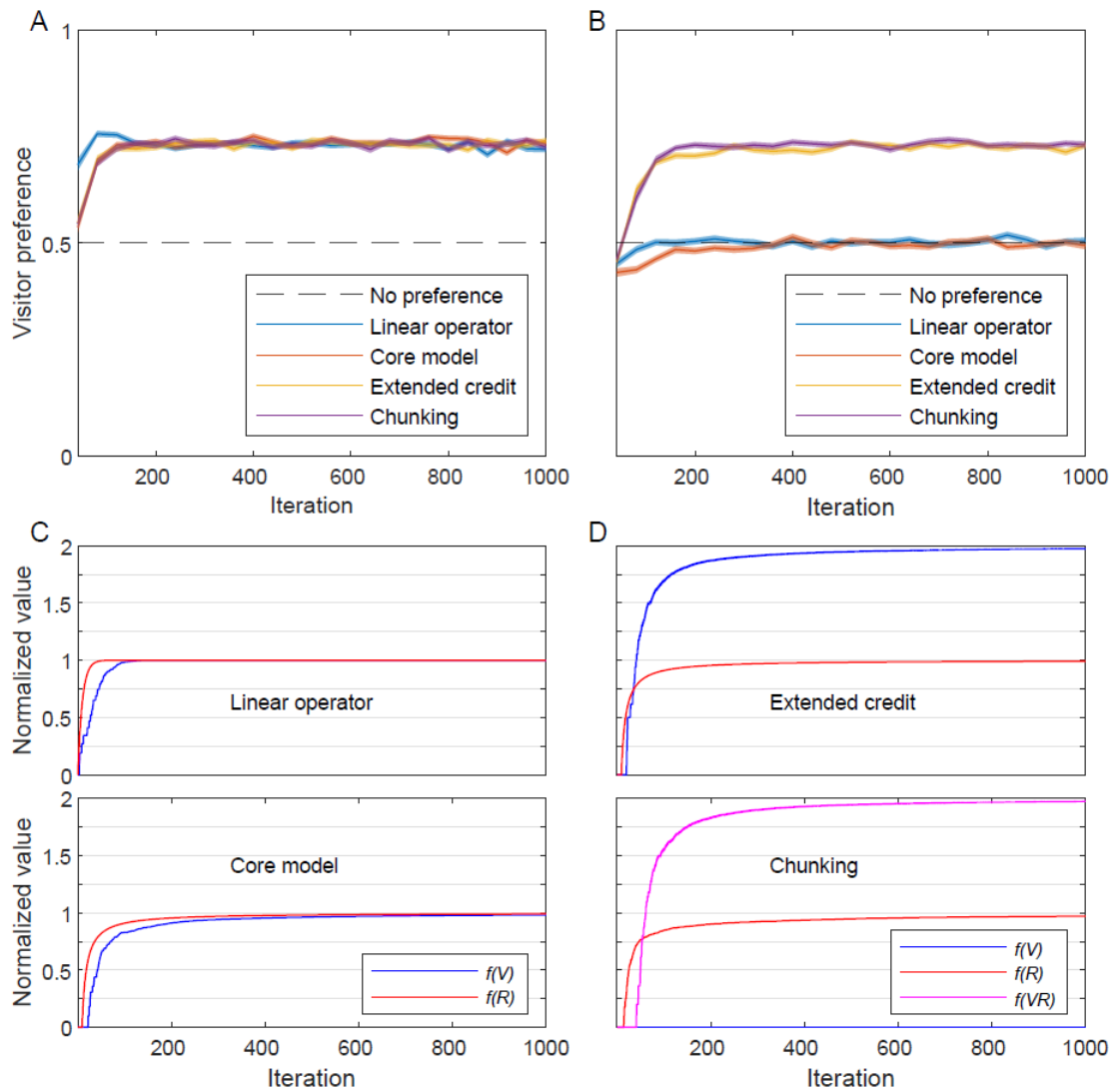
Figure 3. Simulating laboratory environments. Four types of learners are compared in the *basic two-choice task* (**A**) and the *laboratory market problem* (**B**): blue – A linear operator learner ($\alpha=0.1$; see text); orange – the core model; yellow – the extended credit model; purple – the chunking model (with $C_p=2$); black dashed-line – the expected choices with no preference (0.5). The preference towards a visitor client, measured as the proportion of choosing a visitor out of all visitor-resident encounters, is plotted as a function of time (iterations), in bins of 40 trials. Both laboratory environments were simulated using 1000 feeding trials (with an empty trial after each feeding trial). The plots depict the mean of 100 simulations for each learner (shades – standard error of the mean). **C)** The value of the different cues as perceived by the non-successful models, the linear operator (top) and the core model (bottom), in a single simulation of the *laboratory market problem*: blue – V; red – R. **D)** The values perceived by the successful models, the extended-credit model (top) and the chunking model (bottom) in a single simulation of the *laboratory market problem*: blue – V; red – R; magenta – VR. Note that the chunking model, in this task, quickly creates the VR chunk, even before any value is attached to V itself.

310

311

12

The natural market problem
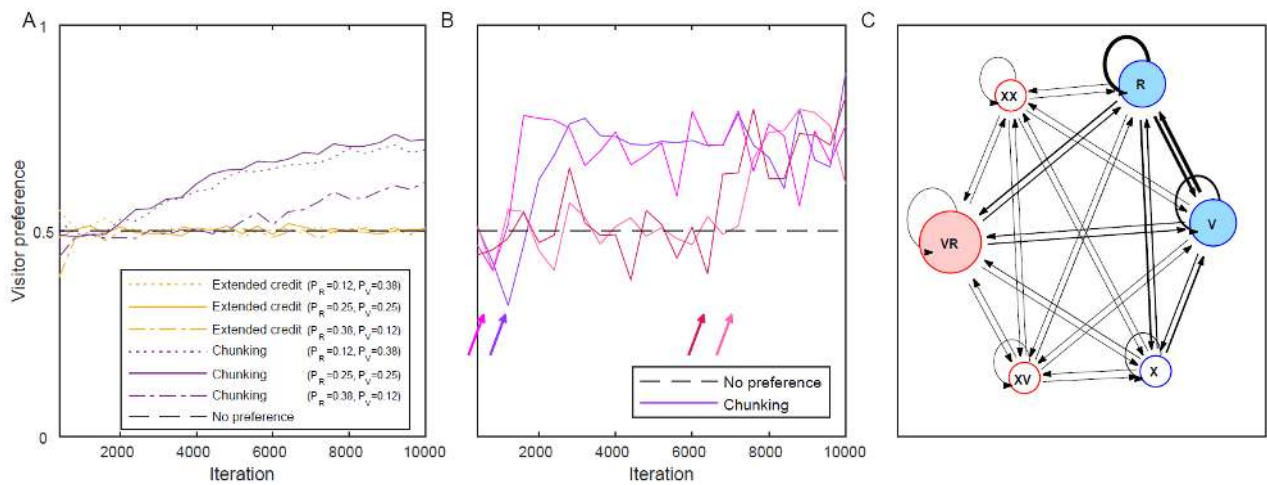
312   **The natural market problem**

313   For the natural market problem, we only present the results of the learning models that were successful

314   in solving the laboratory market problem (as expected, the core and the linear-operator models that

315   failed to solve the laboratory problem also fail to solve the more complex natural problem, data not

316   shown).

317   The extended-credit model that was sufficient for solving the laboratory market problem failed to solve

318   the natural market problem (i.e., to prefer a visitor client) regardless of the overall client density or the

319   relative frequencies of different clients (see examples in Fig. 4A, yellow lines). The reason for that is

320   that in the *natural market problem* all pair sequences can occasionally appear, including a resident after

321   a resident (and thus R is credited with 2 food items), a visitor after a visitor (and thus V is credited with

322   1), and even a resident and then a visitor (when there is no empty trial after serving the resident, which

323   again credit R with 2). Thus, assigning credit for a cue for the value of the next cue causes the differences

324   between $f(R)$ and $f(V)$ to vanish. Still, the sequence $V \rightarrow R$ may occur more often than the sequence

325    $R \rightarrow V$ (at least as long as the cleaner do not prefer R), since whenever the two types of clients appear

326   simultaneously, $V \rightarrow R$ occurs if the cleaner chooses to serve the visitor first, and $R \rightarrow V$ occurs only

327   when a visitor appears by chance in a new trial after the cleaner has served a resident. As a result, $f(V)$

328   might be slightly greater than $f(R)$ in some situations. However, in order to respond to such slight

329   differences, the model's soft-max decision rule should be 'hardened' (become more similar to a

330   maximum-based rule). This would suppress exploration and make the model always choose the most

331   frequent client type (as its value increases faster), which is the resident in most cases since the visitor

332   leaves if not served.

333   In contrast to the extended credit model, the chunking model solved the *natural market problem*

334   successfully in a wide range of client distributions (Fig. 4A, purple lines). To solve this task, the

335   chunking model only needs to create the chunk VR, which in turn imposes a preference for the visitor,

336   as VR is always associated with two units of food reward. The time of creating the VR chunk may vary

337   according to the stochastic order of the trials experienced by each individual (see examples in Fig 4B).

338   But on average, as the simulation advances, the chances of a cleaner using the chunking model to create

339   the VR chunk and thus to choose a visitor increases (Fig. 4A, purple lines). Figure 4C depicts an example

340   of the internal representation of the chunking model at the end of a simulation of the *natural market*

341   *problem*. Note that the chunking model creates chunks regardless of the reward, and depending only on

342   the statistics of cue occurrence. Thus, it may also create chunks containing the cue for an empty arena

343   (X), or for other various combinations (e.g. XR, XV, RR, VV, etc.). In most cases these chunks do not

344   influence the cleaner's decisions as they represent states that require no choice (see Fig. 2). However,

345   as we shall see below, in the natural setting there is also a risk of creating the RV chunk (rather than VR

346   chunk), which can bias the cleaner's decision, implying that chunking should be limited to avoid over-

347   chunking.

Figure 4. Simulations of the *natural market problem*. A) The preference for a visitor by the extended-credit model (yellow) and the chunking model (purple) are presented for client density (*1-P$_0$*) of 0.5 and for three different distributions of client types: *P$_R$*=0.12 and *P$_V$*=0.38 (dotted lines), *P$_R$*=0.25 and *P$_V$*=0.25 (solid lines), *P$_R$*=0.38 and *P$_V$*=0.12 (dotted-dashed lines). Black dashed line – no preference (0.5). *C$_p$*=2 (for the chunking model). **B)** Four simulations of the chunking model in the *natural market problem* (with *P$_R$*=0.25 and *P$_V$*=0.25). Note how the preference towards a visitor sharply increases after the creation of the *VR* chunk (depicted with an arrow of a corresponding colour for each simulation). **C)** The internal representation of the chunking model at the end of a simulation as in **(B)**. Blue – basic (initial) elements, red – chunk elements, filled nodes – the relevant elements for the decision process. The size of the circle is relative to the value (association with food reward) of the element. The width of the directed edges (black arrows) is relative to the weight (*W*) of the transitions between elements.

348

The fine-tuning of chunking behaviour and the effect of ecological conditions

350   The behaviour of the chunking model is controlled by the chunking avoidance parameter *C$_p$* (Eq. 2).

351   Large values of *C$_p$* prevent any chunking and the model is reduced back to the core model (which do

352   not develop a preference towards the optimal choice). On the other side, too low values of *C$_p$* cause

353   'over-chunking'. Therefore, the optimal value of *C$_p$* will depend on the ecological conditions: the

354   overall client density, and the frequency of the different client types. If there are many clients per

355   cleaner, cleaners will often be solicited. Therefore, a visitor may regularly appear right after a resident

356   – not because the visitor waits for service, but because a new visitor client enters into the arena by

357   chance. Thus, there is a risk that the misleading chunk *RV* might be created, as well as the beneficial

358   chunk *VR*. The reason we view the *RV* chunk as misleading is that faced with a choice between a visitor

359   and a resident, the cleaner can now consider both sequences of actions, *V → R* and *R → V*, and choose

360   between them according to their expected values. Although the value of the chunk *VR*, *f(VR)*, would

361   approach 2 and hence be higher than the value of the chunk *RV* (with *f(RV)* lower than 2), the decision

362   rule allows some proportion of choosing the *RV* chunk (exploration), which result in serving the resident

363   first. In other words, over-chunking reduces the strength of the preference for the optimal choice. The

364   balance between under-chunking and over-chunking implies the existence of optimal C_p values

365   (balancing between the need to create the *VR* chunk but not the *RV* chunk). Importantly, these optimal

366   C_p values depend on two ecological conditions: the overall client density, and the frequency of the

367   different client types, which determine how frequently the sequences $V \rightarrow R$ and $R \rightarrow V$ are likely to

368   be encountered. The effect of these ecological conditions on C_p and on the success of solving the *natural*

369   *market problem* is shown in Fig. 5. It can be seen that in some extreme ecological conditions (of high

370   client densities) it would be difficult for a cleaner fish using the chunking model to solve the market

371   problem with any C_p (Fig. 5A, black dots, and Fig. 5B, blue shades representing low preference for

372   visitors), since empty spots are rare events and most choices of the chunk *RV* result in obtaining two

373   units of food (from the resident and the subsequent served client from the next trial). Fortunately for

374   the cleaners, solving the market problem under these high client density conditions is not important in

375   nature as high client densities lead to near permanent demand for cleaning. Yet, in most simulated

376   ecological conditions where solving the market problem is important, an optimal C_p value (Fig. 5A)

377   that induced a preference towards a visitor (Fig. 5B) was found.
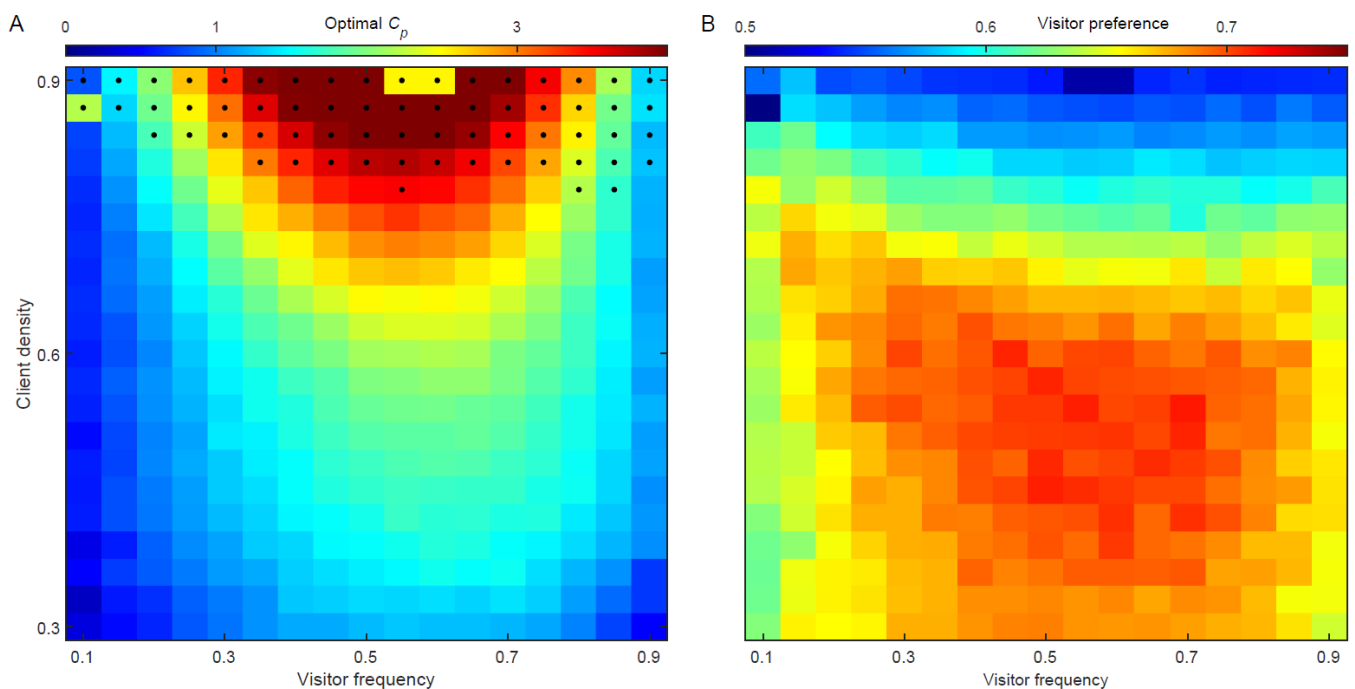
Figure 5. The link between ecological conditions, optimal C_p, and the success of the chunking model in the *natural market problem*. A) Optimal C_p values (that provide the strongest preference towards a visitor), indicated by color, as a function of two ecological conditions: the visitor frequency, $\frac{P_V}{P_V + P_R}$ (the fraction of visitors out of all clients), and the overall client density, $1 - P_0$. The C_p values were estimated by running the simulations with 1000 values equally distributed between 0 and 5, fitting a Gaussian to the resulting visitor's preferences, and finding its peak. Black dots depict conditions in which even the optimal C_p

values resulted in a preference of less than 0.6 towards the visitor client. **B)** The preference (colour) towards the visitor client when the optimal C_p values are used in different ecological conditions.

378

379   To visualize the importance of the over-chunking problem, Figure S1 presents the frequency of
380   appearances of each possible chunk (among 100 simulations) in four different ecological conditions,
381   showing that when the chances of generating both the *VR* and *RV* chunks are similar (Fig. S1B), the
382   preference towards visitors vanishes (compare with the relevant point of 0.5 visitor frequency and 0.9
383   client density in Fig. 5B).

384   Finally, our simulations show a significant positive correlation (linear regression: R^2=0.78, p<0.001)
385   between the frequency of simultaneous arrival of a visitor and a resident to the arena (hereafter: *r+v*;
386   Fig. 6A) and the optimal C_p value (Fig. 6B). That is, when the combination of client density and visitors'
387   relative frequency increases the frequency of *r+v* pairs, a higher value of C_p should be used by the
388   cleaners in order to increase the threshold of statistical significance allowing a chunk to be created. In
389   contrast, when *r+v* pairs are rare, the probability of creating the misleading chunk *RV* is low so that
390   lowering C_p is adaptive: it increases the likelihood of creating the beneficial chunk *VR* with almost no
391   risk of creating the misleading chunk *RV*, which allows a strong preference for visitors to develop. Note,
392   that when the frequency of *r+v* pairs is especially high (above 0.3; Fig. 6), there appears to be no C_p
393   value that could balance between over- and under-chunking and the preference for visitors goes below
394   0.6 (only black dots appear at this range in Fig. 6B). More generally, a tendency to chunk too soon (e.g.,
395    $C_p = 0.5$) or too late (e.g., $C_p = 2.5$) resulted in poor performance under most combinations of client
396   densities and visitor frequencies (Fig. S2).

397

398
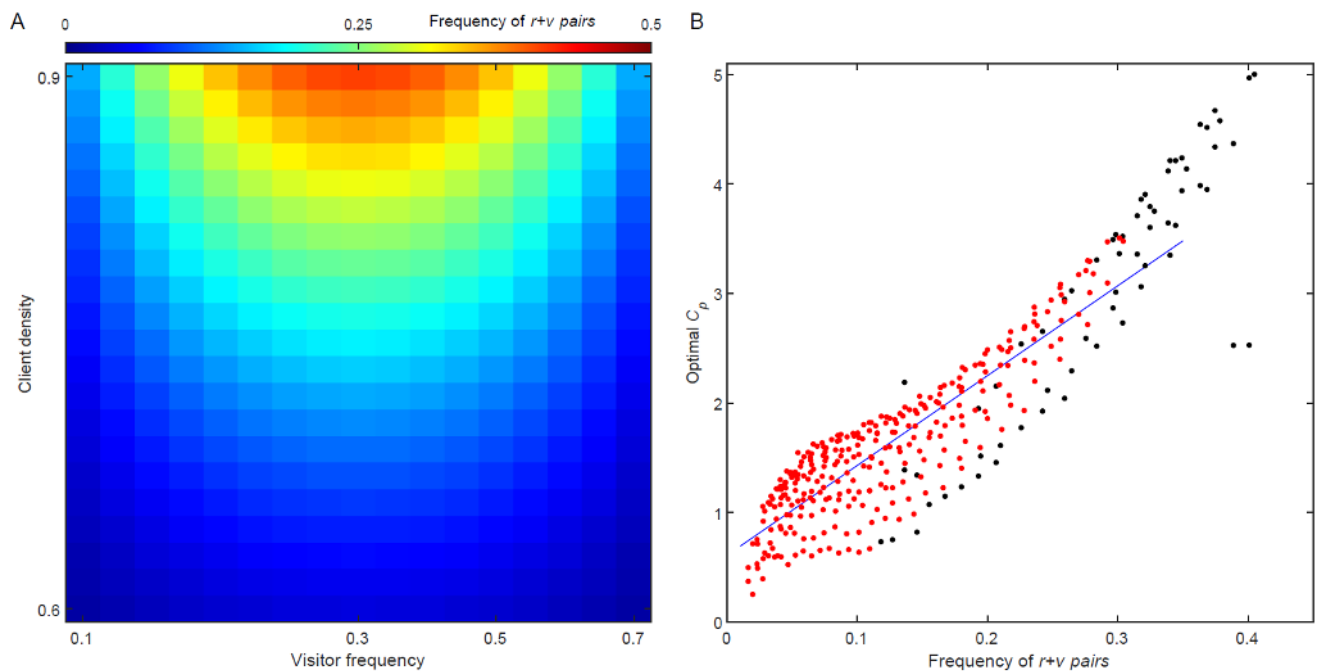
Figure 6. Correlation between Optimal C_p values and the frequency of resident and visitor pairs. A) The frequency of simultaneous appearance of resident and visitor ($r+v$ pairs) in the arena, indicated by colour, out of all simulation trials (including empty and half empty trials) in the *natural market problem*. These are not stochastic values, but a feature of the simulated environment. **B)** The optimal C_p value (as in Fig 5A) as a function of the frequency of $r+v$ pairs. Black dots – values obtained from simulations that achieved a preference towards a visitor lower than 0.6 (corresponding to the black dots in Fig. 5A). Blue line – linear regression of the optimal C_p values which achieved successful solutions (red dots; $R^2 =$ 0.78).

399

400

Discussion

401

402   Chunking mechanisms are essential to represent structured data in the brain and have probably played

403   a pivotal role in the evolution of cognition [28,30,33,41,42]. Yet, a possible challenge in the evolution

404   of chunking is that incorrect chunking and over-chunking may lead to maladaptive behaviours and to

405   cognitive impairments [66,67]. Indeed, the problem of under- or over-chunking arises whenever sensory

406   input has to be chunked or segmented (reviewed in [33]). Normally, the problem is difficult to track

407   because incoming data can be chunked in multiple ways and the number of possible chunks grows

408   exponentially with the amount of data. This problem is well appreciated, for example, in the case of

409   word segmentation during language learning in humans [32,68], or in the representation of behavioural

410   sequences by animals [69]. Our analyses show that the market problem solved by cleaner fish in the

411 wild offers a relatively simple model system to study the evolution of chunking. It is not only simple

412 and tractable, but it involves a case where the function of chunking and its fitness consequences are

413 well understood and are ecologically relevant, the adaptive and maladaptive chunks can be clearly

414 identified (i.e., *VR* versus *RV*), and it can be studied experimentally and in relation to variable ecological

415 conditions (e.g., [44,50,51]).

416 We implemented this approach by placing the same general problem of making a decision that doubles

417 food intake in different sequential contexts that cleaners face in the wild. We show how solutions

418 depend on increasingly complex learning rules. A simple two-choice task can be solved with basic

419 reinforcement learning models such as the linear-operator or our equivalent core model. A more

420 challenging task where doubling the amount of food is consistently due to consequences of an initial

421 choice (i.e., the *laboratory market task*) requires an extended-credit learning model that picks up a

422 consistent chain of events. Finally, if cleaners face diverse sequences of events, as in the *natural market*

423 *problem*, relevant causal chains of subunits that lead to doubling the food intake must be identified and

424 chunked so that the animal can optimise food intake.

425 We also demonstrate that when facing diverse sequences of events, having the ability to chunk may not

426 be sufficient. It is critical that the decision to create chunks, captured by the chunking parameter C_p, be

427 adjusted to ecological conditions. Moreover, our simulations also show that under some extreme

428 conditions, even the optimal chunking parameter may not be sufficient for developing a preference for

429 the ephemeral reward. In the cleaners' market problem, it happens when the probability of encountering

430 the sequences of the useful and misleading chunks, *VR* and *RV*, respectively, is so similar that no

431 chunking parameter can allow the creation of *VR* while preventing the creation of *RV*. As mentioned

432 earlier, in the case of the cleaner fish, this may not be a problem because it happens under conditions

433 of high client densities where preferring the ephemeral reward (i.e. visitors) is not necessary. It is yet to

434 be studied how common are such conditions in other problems animals face in nature, and to what

435 extent using the right chunking parameter is sufficient for successfully balancing the trade-off between

436 under- and over-chunking.

437 Demonstrating the trade-off between adaptive chunking and over-chunking yields a new perspective on

438 the cognitive basis of cleaner fish 'cleverness' in their choices of clients. Solving the natural market

439 problem does not represent an "all or none" cognitive ability but rather the ability to correctly adjust a

440 more basic cognitive ability, which is the ability to create chunks. As it stands, many animals are capable

441 of creating chunks and configurations in their memory representation (see Introduction), but only those

442 applying the chunking parameters suitable to the required conditions will solve the natural market

443 problem. The trade-off between chunking and over-chunking may also explain why chunking (and

444 configurational learning) takes time and may thus be viewed as difficult. Our model suggests that there

445 is nothing really difficult in creating chunks quickly but that the process of chunking evolved to be slow

446 in order to prevent over-chunking. Note that the idea that learning may evolve to be slow as a result of

447   a trade-off is not new. It is implied in the optimization of learning rate parameters to balance between

448   exploration and exploitation in reinforcement learning models [70,71], and was also suggested as a way

449   to minimize recognition errors [72,73].

450

451   **The mechanism of chunking**

452   Our chunking model specifies the statistical conditions required for the formation of chunks and

453   describes how chunks are represented in the network (Eqs. 2 and 3; Fig. 1B). Yet, it does not explain

454   how chunks are actually created. In other words, it does not explain how it happens that under the

455   conditions specified by Equations 2 and 3, a chunk in the network suddenly appears. While the neuronal

456   coding of such information is still poorly understood [29], a fairly explicit implementation of the process

457   of chunk formation using neuronal-like processes may be possible. We can think of the required number

458   of co-occurrences of V and R that is represented by the left side of Eq. 2 as the weight of their associative

459   strength. Accordingly, a chunk representing the sequence VR is created when the weight of the edge

460   leading from V to R passes a certain threshold. The formation of a chunk may be a result of another

461   node in the network that receives signals from both neuronal units (or more precisely, from R soon after

462   V), and thus increases in weight and becomes the "chunk node" representing the repeated occurrences

463   of the sequence VR (as in Fig. 1B). The threshold weight required for the creation of a chunk can thus

464   act as the chunking parameter C_p in our model and be optimized in line with Eq. 2.

465   In our model, that was kept as simple as possible, we assumed that weight increases by one unit per

466   observation and does not decay over time. Realistically, however, different combinations of weight

467   adjustment rates determine the timing crossing the threshold for chunk formation. For example, slow

468   increase in weight with a relatively fast decay require frequent co-occurrences in order to reach the

469   threshold, creating a test for the chunk's statistical significance [19,33,41]. Thus, the chunking

470   parameter in our model can be implemented by several mechanisms. We can hence view this parameter

471   (or parameters) more generally as those effecting the tendency to form chunks (or the tendency to use

472   configurational rather than elemental learning).

473   The optimization of the chunking parameters to ecological conditions may occur over generations

474   through selection acting directly on parameter values, or instead (or in addition) cleaners may have

475   evolved phenotypic plasticity with respect to the chunking parameter. For example, a rule instructing

476   the cleaner to vary (loosen) the chunking parameters (i.e., explore) when in poor conditions and stop

477   altering it (fasten) when in good conditions (i.e., exploit) may bring the chunking parameters to get

478   fixated around the values associated with best performance. Another possibility is that cases where a

479   visitor is leaving without waiting are experienced by the cleaner as aversive (a loss of a meal) and the

480   aversive saliency of such events has evolved to reduce the chunking threshold (which increases the

481   likelihood of chunking when solving the market problem is indeed necessary).

19

482

Implications of our results on the interpretation of empirical studies

483

484   A major insight from our model in comparison to Quiñones et al. [62] is that animals only need the
485   ability to detect chains of events (rather than chunking) in order to solve the laboratory market problem.
486   Accordingly, it is not at all clear that differences between species in performance in the laboratory
487   market task are due to different chunking abilities or different values of chunking parameters. It is hence
488   important to use a more complex design of the market task (which resemble the natural setting for which
489   chunking is necessary) on species that have solved some form of the laboratory task, i.e. cleaner fish,
490   African grey parrots and capuchin monkeys [54,59]. Truskanov et al. [74] designed such a task,
491   exposing cleaner fish to 50% of presentations of visitor and resident ($r+v$) plates as well as to 25% $r+r$
492   and 25% $v+v$ presentations. While a few cleaners solved this task, overall performance tended to be
493   lower than in the standard laboratory market task. Applying our learning models to this non-standard
494   (complex) market task showed that the extended credit model yields at best a slight preference for
495   visitors, while the chunking model yields high performance (see Supplementary Information and Fig.
496   S3). The study by Truskanov et al. thus yields experimental evidence that (some) cleaner fish can chunk.
497   The task could also be adapted to test whether imposing "early commitment" that helped pigeons in
498   solving the standard laboratory market problem [60] can also help to solve the natural problem, for
499   which chunking ability is needed. Alternatively, "early commitment" can only help in extending the
500   credit given to the initial choice (to the second reward as well as to the first one), which can solve the
501   laboratory market problem but not the natural one.

502   Based on our model and simulations, there are currently multiple ways to explain the documented
503   intraspecific variation in cleaner fish performance in both the standard and the complex laboratory
504   market tasks [46,48,51,74]. First, variation in the laboratory market task may be related to whether
505   individuals solve the problem by chunking or by chaining (extended credit) mechanisms, and to
506   individual variation in the fine-tuning of the parameters of each mechanism. Second, assuming that
507   cleaners use chunking to solve the tasks, variation in their performance may be attributed to some
508   limitations or time lags in optimizing the chunking parameters to current conditions in the field, or to
509   the specific conditions in the lab. Such limitations and time lags are expected for both genetic and
510   phenotypically plastic adjustments because in the cleaners' natural habitat, client densities and visitor
511   frequencies can vary greatly across years and microhabitats [48,51], causing both inter- and intra-
512   individual variation within individual lifetimes.

513   Importantly, these interpretations make related assumptions amenable for future testing. For example,
514   that fast-solving cleaners use chunking even in the laboratory market task even though chaining would
515   suffice, and that cleaners apply their field experience and developed C_p value to the lab task. Some
516   empirical results are already in line with the second assumption. First, the best predictor of high cleaner

517 performance in the laboratory task is high cleaner fish density [46], which in terms of our model implies

518 low client density (per individual cleaner) and therefore low optimal C_p that promotes faster chunking

519 (see Fig. 5A and Fig. 6). Second, individuals with relatively larger forebrains are more likely to be

520 found in areas where they frequently face the market problem; [75]. Third, on a local scale, individuals

521 with relatively larger forebrains performed according to what appears to be the locally best strategy: to

522 solve the task if living in a high cleaner density area, and to fail the task if living in a low-cleaner density

523 area [76]. In terms of our model, such high and low cleaner densities correspond to relatively low and

524 high client densities that favour low and high C_p values, respectively (see Fig. 5A). Bringing such

525 cleaners to lab implies that those who developed low C_p in their natural habitat are more likely to pass

526 the test than those who developed high C_p, which may explain Triki et al.'s results [76].

527

528 **Conclusions and implications for the study of advanced cognitive abilities**

529 The cleaner fish ability to solve the market problem has presumably evolved on the background of its

530 unique ecology and may be rightfully viewed as a surprisingly advanced cognitive ability for a (small

531 brain) fish. However, by modeling the learning mechanisms required for this remarkable ability, we

532 tried to put the cleaner fish story within the broader context of cognitive evolution, viewing it as a

533 potential model for the evolution of chunking mechanisms. While the importance of chunking is usually

534 considered within cognitive systems that are already highly advanced, the simple setting of the market

535 problem allowed us to explicitly analyze the process of chunk formation, elucidating the trade-off

536 between creating useful and misleading chunks, and demonstrating the importance of adjusting the

537 chunking parameters to ecological conditions. We hope that the approach taken here could eventually

538 be applied in the study of other cognitive abilities, identifying the learning mechanisms and the fine-

539 tuning of their parameters required for their success, and mapping them not only along phylogenetic

540 trees but also along evolutionary axes of explicit incremental changes in learning and cognitive

541 mechanisms.

542

543

544 # Competing interests

545 The other authors declare that there are no competing interests.

546

547

548

References

1. Bluff LA, Weir AAS, Rutz C, Wimpenny JH, Kacelnik A. 2007 Tool-related cognition in New Caledonian crows. *Comp. Cogn. Behav. Rev.* **2**, 1–25.

2. Emery NJ, Clayton NS. 2004 The Mentality of Crows: Convergent Evolution of Intelligence in Corvids and Apes. *Science* **306**, 1903 LP – 1907. (doi:10.1126/science.1098410)

3. Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. 2006 Recursive syntactic pattern learning by songbirds. *Nature* **440**, 1204–1207. (doi:10.1038/nature04675)

4. Scarf D, Boy K, Uber Reinert A, Devine J, Güntürkün O, Colombo M. 2016 Orthographic processing in pigeons (Columba livia). *Proc. Natl. Acad. Sci.* **113**, 11272 LP – 11276. (doi:10.1073/pnas.1607870113)

5. Murphy RA, Mondragón E, Murphy VA. 2008 Rule Learning by Rats. *Science* **319**, 1849 LP – 1851. (doi:10.1126/science.1151564)

6. Harten L, Katz A, Goldshtein A, Handel M, Yovel Y. 2020 The ontogeny of a mammalian cognitive map in the real world. *Science* **369**, 194 LP – 197. (doi:10.1126/science.aay3354)

7. MacLean EL *et al.* 2012 How does cognition evolve? Phylogenetic comparative psychology. *Anim. Cogn.* **15**, 223–238. (doi:10.1007/s10071-011-0448-8)

8. Cauchoix M, Chaine AS. 2016 How Can We Study the Evolution of Animal Minds? *Front. Psychol.* **7**, 358.

9. Pinker S, Bloom P. 1990 Natural language and natural selection. *Behav. Brain Sci.* **13**, 707–727. (doi:10.1017/S0140525X00081061)

10. Berwick RC, Friederici AD, Chomsky N, Bolhuis JJ. 2013 Evolution, brain, and the nature of language. *Trends Cogn. Sci.* **17**, 89–98.

11. Whiten A, editor. 1991 Natural theories of mind: Evolution, development and simulation of everyday mindreading. *Nat. Theor. mind Evol. Dev. Simul. everyday mindreading.* , xiv, 362–xiv, 362.

12. Gallese V, Goldman A. 1998 Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* **2**, 493–501. (doi:https://doi.org/10.1016/S1364-6613(98)01262-5)

13. Rizzolatti G, Craighero L. 2004 The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169–192. (doi:10.1146/annurev.neuro.27.070203.144230)

14. Solan Z, Horn D, Ruppin E, Edelman S. 2005 Unsupervised learning of natural languages. *Proc. Natl. Acad. Sci.* **102**, 11629–11634. (doi:10.1073/PNAS.0409746102)

580   15.   Heyes C. 2010 Where do mirror neurons come from? *Neurosci. Biobehav. Rev.* **34**, 575–583.
581   (doi:https://doi.org/10.1016/j.neubiorev.2009.11.007)

582   16.   Heyes CM. 1998 Theory of mind in nonhuman primates. *Behav. Brain Sci.* **21**, 101–114.
583   (doi:DOI: 10.1017/S0140525X98000703)

584   17.   Galef BG. 1992 The question of animal culture. *Hum. Nat.* **3**, 157–178.
585   (doi:10.1007/BF02692251)

586   18.   Lipkind D *et al.* 2013 Stepwise acquisition of vocal combinatorial capacity in songbirds and
587   human infants. *Nature* **498**, 104–108.

588   19.   Kolodny O, Edelman S, Lotem A. 2015 Evolution of protolinguistic abilities as a by-product
589   of learning to forage in structured environments. *Proc. R. Soc. B Biol. Sci.* **282**, 20150353.
590   (doi:10.1098/rspb.2015.0353)

591   20.   Kolodny O, Lotem A, Edelman S. 2015 Learning a Generative Probabilistic Grammar of
592   Experience: A Process-Level Model of Language Acquisition. *Cogn. Sci.* **39**, 227–267.
593   (doi:https://doi.org/10.1111/cogs.12140)

594   21.   Galef BG. 2015 Laboratory studies of imitation/field studies of tradition: Towards a synthesis
595   in animal social learning. *Behav. Processes* **112**, 114–119.
596   (doi:https://doi.org/10.1016/j.beproc.2014.07.008)

597   22.   Bellmund JLS, Gärdenfors P, Moser EI, Doeller CF. 2018 Navigating cognition: Spatial codes
598   for human thinking. *Science* **362**, eaat6766. (doi:10.1126/science.aat6766)

599   23.   Skinner BF. 1936 The Reinforcing Effect of a Differentiating Stimulus. *J. Gen. Psychol.* **14**,
600   263–278. (doi:10.1080/00221309.1936.9713154)

601   24.   Williams BA. 1994 Conditioned Reinforcement: Experimental and Theoretical Issues. *Behav.*
602   *Anal.* **17**, 261–285. (doi:10.1007/BF03392675)

603   25.   Kolodny O, Edelman S, Lotem A. 2014 The evolution of continuous learning of the structure
604   of the environment. *J. R. Soc. Interface* **11**, 20131091–20131091. (doi:10.1098/rsif.2013.1091)

605   26.   Enquist M, Lind J, Ghirlanda S. 2019 The power of associative learning and the ontogeny of
606   optimal behaviour. *R. Soc. Open Sci.* **3**, 160734. (doi:10.1098/rsos.160734)

607   27.   Leadbeater E. 2015 What evolves in the evolution of social learning? *J. Zool.* **295**, 4–11.
608   (doi:https://doi.org/10.1111/jzo.12197)

609   28.   Pearce JM. 2002 Evaluation and development of a connectionist theory of configural learning.
610   *Anim. Learn. Behav.* **30**, 73–95. (doi:10.3758/BF03192911)

611   29.   Duncan K, Doll BB, Daw ND, Shohamy D. 2018 More Than the Sum of Its Parts: A Role for

612   the Hippocampus in Configural Reinforcement Learning. *Neuron* **98**, 645-657.e6.

613   (doi:10.1016/j.neuron.2018.03.042)

614   30.   Gobet F, Lane P, Croker S, Cheng P, Jones G, Oliver I, Pine J. 2001 Chunking mechanisms in

615   human learning. *Trends Cogn. Sci.* **5**, 236–243.

616   31.   Jones G. 2012 Why chunking should be considered as an explanation for developmental

617   change before short-term memory capacity and processing speed. *Front. Psychol.* **3**, 1–8.

618   (doi:10.3389/fpsyg.2012.00167)

619   32.   Brent MR. 1999 Speech segmentation and word discovery: A computational perspective.

620   *Trends Cogn. Sci.* **3**, 294–301. (doi:10.1016/S1364-6613(99)01350-9)

621   33.   Goldstein MH, Waterfall HR, Lotem A, Halpern JY, Schwade J a., Onnis L, Edelman S. 2010

622   General cognitive principles for learning structure in time and space. *Trends Cogn. Sci.* **14**,

623   249–258. (doi:10.1016/j.tics.2010.02.004)

624   34.   Whitlow JW, Wagner AR. 1972 Negative patterning in classical conditioning: Summation of

625   response tendencies to isolable and configurai components. *Psychon. Sci.* **27**, 299–301.

626   (doi:10.3758/BF03328970)

627   35.   Rescorla RA, Wagner AR. 1972 A theory of pavlovian conditioning: variations in the

628   effectiveness of reinforcement and nonreinforcement. In *Clasical conditioning II: current*

629   *research and theory* (eds AH Black, WF Prokasy), pp. 64–99. Appleton-Century-Crofts.

630   36.   Mackintosh NJ. 1975 A theory of attention: Variations in the associability of stimuli with

631   reinforcement. *Psychol. Rev.* **82**, 276–298. (doi:10.1037/h0076778)

632   37.   Sutherland RJ, Rudy JW. 1989 Configural association theory: The role of the hippocampal

633   formation in learning, memory, and amnesia. *Psychobiology* **17**, 129–144.

634   (doi:10.3758/BF03337828)

635   38.   Papadimitriou A, Wynne CDL. 1999 Preserved negative patterning and impaired spatial

636   learning in pigeons (Columba livia) with lesions of the hippocampus. *Behav. Neurosci.* **113**,

637   683–690. (doi:10.1037/0735-7044.113.4.683)

638   39.   Fournier DI, Todd TP, Bucci DJ. 2019 Permanent damage or temporary silencing of

639   retrosplenial cortex impairs the expression of a negative patterning discrimination. *Neurobiol.*

640   *Learn. Mem.* **163**, 107033. (doi:https://doi.org/10.1016/j.nlm.2019.107033)

641   40.   Conway CM. 2020 How does the brain learn environmental structure? Ten core principles for

642   understanding the neurocognitive mechanisms of statistical learning. *Neurosci. Biobehav. Rev.*

643   **112**, 279–299. (doi:10.1016/j.neubiorev.2020.01.032)

644   41.   Lotem A, Halpern JY. 2012 Coevolution of learning and data-acquisition mechanisms: a

645   model for cognitive evolution. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 2686–2694.

646   (doi:10.1098/rstb.2012.0213)

647 42. Lotem A, Halpern JY, Edelman S, Kolodny O. 2017 The evolution of cognitive mechanisms in

648   response to cultural innovations. *Proc. Natl. Acad. Sci.* **114**, 7915 LP – 7922.

649   (doi:10.1073/pnas.1620742114)

650 43. Côté IM. 2000 Evolution and ecology of cleaning symbioses in the sea. In *Oceanography and*

651   *Marine Biology: An Annual Review: Volume 38: An Annual Review* (eds RN Gibson, M

652   Barnes), pp. 311–355. CRC Press.

653 44. Bshary R. 2001 The cleaner fish market. In *Economics in Nature* (eds R Noë, JARAM Hooff,

654   P Hammerstein), pp. 146–172. Cambridge University Press.

655 45. Bshary R, Noë R. 2003 Biological markets: the ubiquitous influence of partner choice on the

656   dynamics of cleaner fish-client reef fish interactions. In *Genetic and cultural evolution of*

657   *cooperation* (ed P Hammerstein), pp. 167–184. MIT press.

658 46. Triki Z, Wismer S, Rey O, Ann Binning S, Levorato E, Bshary R. 2019 Biological market

659   effects predict cleaner fish strategic sophistication. *Behav. Ecol.* **30**, 1548–1557.

660   (doi:10.1093/beheco/arz111)

661 47. Bshary R. 2011 Machiavellian Intelligence in Fishes. In *Fish Cognition and Behavior*, pp.

662   277–297. (doi:https://doi.org/10.1002/9781444342536.ch13)

663 48. Wismer S, Pinto AI, Vail AL, Grutter AS, Bshary R. 2014 Variation in Cleaner Wrasse

664   Cooperation and Cognition: Influence of the Developmental Environment? *Ethology* **120**,

665   519–531. (doi:https://doi.org/10.1111/eth.12223)

666 49. Bshary R, Grutter AS. 2002 Experimental evidence that partner choice is a driving force in the

667   payoff distribution among cooperators or mutualists: the cleaner fish case. *Ecol. Lett.* **5**, 130–

668   136. (doi:https://doi.org/10.1046/j.1461-0248.2002.00295.x)

669 50. Truskanov N, Emery Y, Bshary R. 2020 Juvenile cleaner fish can socially learn the

670   consequences of cheating. *Nat. Commun.* **11**, 1159. (doi:10.1038/s41467-020-14712-3)

671 51. Triki Z, Wismer S, Levorato E, Bshary R. 2018 A decrease in the abundance and strategic

672   sophistication of cleaner fish after environmental perturbations. *Glob. Chang. Biol.* **24**, 481–

673   489. (doi:10.1111/gcb.13943)

674 52. Salwiczek LH *et al.* 2012 Adult Cleaner Wrasse Outperform Capuchin Monkeys,

675   Chimpanzees and Orang-utans in a Complex Foraging Task Derived from Cleaner – Client

676   Reef Fish Cooperation. *PLoS One* **7**, e49068.

677 53. Staddon JER, Cerutti DT. 2003 Operant Conditioning. *Annu. Rev. Psychol.* **54**, 115–144.

678   (doi:10.1146/annurev.psych.54.101601.145124)

679 54. Prétôt L, Bshary R, Brosnan SF. 2016 Factors influencing the different performance of fish
680   and primates on a dichotomous choice task. *Anim. Behav.* **119**, 189–199.
681   (doi:https://doi.org/10.1016/j.anbehav.2016.06.023)

682 55. Zentall TR, Case JP, Luong J. 2016 Pigeon's (Columba livia) paradoxical preference for the
683   suboptimal alternative in a complex foraging task. *J. Comp. Psychol.* **130**, 138–144.

684 56. Zentall TR, Case JP, Berry JR. 2017 Rats' acquisition of the ephemeral reward task. *Anim.*
685   *Cogn.* **20**, 419–425. (doi:10.1007/s10071-016-1065-3)

686 57. Zentall TR, Case JP. 2018 The Ephemeral-Reward Task: Optimal Performance Depends on
687   Reducing Impulsive Choice. *Curr. Dir. Psychol. Sci.* **27**, 103–109.
688   (doi:10.1177/0963721417735522)

689 58. Prétôt L, Mickelberg J, Carrigan J, Stoinski T, Bshary R, Brosnan SF. 2020 Comparative
690   performance of orangutans (Pongo spp.), gorillas (Gorilla gorilla gorilla), and drills
691   (Mandrillus leucophaeus), in an ephemeral foraging task. *Am. J. Primatol.* **n/a**, e23212.
692   (doi:https://doi.org/10.1002/ajp.23212)

693 59. Pepperberg IM, Hartsfield LA. 2014 Can Grey parrots (Psittacus erithacus) succeed on a
694   "complex" foraging task failed by nonhuman primates (Pan troglodytes, Pongo abelii, Sapajus
695   apella) but solved by wrasse fish (Labroides dimidiatus)? *J. Comp. Psychol.* **128**, 298–306.
696   (doi:10.1037/a0036205)

697 60. Zentall TR. 2020 The paradoxical performance by different species on the ephemeral reward
698   task. *Learn. Behav.* (doi:10.3758/s13420-020-00429-2)

699 61. Zentall TR, Case JP, Berry JR. 2017 Early commitment facilitates optimal choice by pigeons.
700   *Psychon. Bull. Rev.* **24**, 957–963. (doi:10.3758/s13423-016-1173-8)

701 62. Quiñones AE, Leimar O, Lotem A, Bshary R. 2020 Reinforcement Learning Theory Reveals
702   the Cognitive Requirements for Solving the Cleaner Fish Market Task. *Am. Nat.* **195**, 664–
703   677. (doi:10.1086/707519)

704 63. Bush RR, Mosteller F. 1951 A mathematical model for simple learning. *Psychol. Rev.* **58**,
705   313–323. (doi:10.1037/h0054388)

706 64. McNamara JM, Houston AI. 1987 Memory and the efficient use of information. *J. Theor. Biol.*
707   **125**, 385–395. (doi:https://doi.org/10.1016/S0022-5193(87)80209-6)

708 65. Kelleher RT, Gollub LR. 1962 A review of positive conditioned reinforcement. *J. Exp. Anal.*
709   *Behav.* **5**, 543–597. (doi:10.1901/jeab.1962.5-s543)

710 66. Lotem A, Halpern JY. 2008 A data-acquisition model for learning and cognitive development

711   and its implications for autism.

712   67.   Kolodny O, Edelman S, Lotem A. 2015 Evolved to adapt: A computational approach to animal

713   innovation and creativity. *Curr. Zool.* **61**, 350–368. (doi:10.1093/czoolo/61.2.350)

714   68.   Saffran JR, Aslin RN, Newport EL. 1996 Statistical Learning by 8-Month-Old Infants. *Science*

715   **274**, 1926 LP – 1928. (doi:10.1126/science.274.5294.1926)

716   69.   Byrne RW. 1999 Imitation without intentionality. Using string parsing to copy the

717   organization of behaviour. *Anim. Cogn.* **2**, 63–72. (doi:10.1007/s100710050025)

718   70.   Sutton RS, Barto AG. 2018 *Reinforcement learning: An introduction*. 2nd edn. MIT press.

719   71.   Niv Y, Joel D, Meilijson I, Ruppin E. 2002 Evolution of Reinforcement Learning in Uncertain

720   Environments: A Simple Explanation for Complex Foraging Behaviors. *Adapt. Behav.* **10**, 5–

721   24. (doi:10.1177/1059-712302-010001-01)

722   72.   Lotem A, Nakamura H, Zahavi A. 1995 Constraints on egg discrimination and cuckoo-host co-

723   evolution. *Anim. Behav.* **49**, 1185–1209. (doi:https://doi.org/10.1006/anbe.1995.0152)

724   73.   Rodríguez-Gironés MA, Lotem A. 1999 How to Detect a Cuckoo Egg: A Signal-Detection

725   Theory Model for Recognition and Learning. *Am. Nat.* **153**, 633–648. (doi:10.1086/303198)

726   74.   Truskanov N, Emery Y, Porta S, Bshary R. In press. Extending the challenge of the cleaner

727   fish market problem. *Anim. Behav.*

728   75.   Triki Z, Levorato E, McNeely W, Marshall J, Bshary R. 2019 Population densities predict

729   forebrain size variation in the cleaner fish Labroides dimidiatus. *Proc. R. Soc. B Biol. Sci.* **286**,

730   20192108. (doi:10.1098/rspb.2019.2108)

731   76.   Triki Z, Emery Y, Teles MC, Oliveira RF, Bshary R. 2020 Brain morphology predicts social

732   intelligence in wild cleaner fish. *Nat. Commun.* **11**, 6423. (doi:10.1038/s41467-020-20130-2)

733

734

27

735

A role for ecologically-tuned chunking in the evolution of advanced cognition demonstrated by modelling the cleaner fish market problem

738

Supporting Information

740

Laboratory complex market problem

Following the experiments conducted by Truskanov et al. [74] we simulated another environment of the market problem: the *lab complex market problem*. In this environment the cleaner faces a visitor-resident combination in 0.5 of the feeding trials, a resident-resident combination in 0.25 of the feeding trials, and a visitor-visitor combination in 0.25 of the feeding trials. As in the standard *laboratory market problem*, each feeding trial is followed by an empty trial. In the *lab complex market problem*, the extended-credit model generates only weak preference towards the visitor (Fig. S3, yellow line), yet it still did better than the core model or the linear operator that choose the clients with equal probabilities (Fig. S3, orange and blue lines). The reason for this minor preference is that in the *lab complex market problem* serving a second client after serving a visitor is still more frequent than serving a second client after serving a resident (even before any preference has been developed). This is because serving a visitor is followed by an empty trial only in the case of two visitors being presented (one is served and the other leaves), while serving a resident is followed by an empty trial in all combinations and choices except for a resident followed by another resident (in the visitor-resident choice, choosing a visitor means the resident is served last and choosing a resident also makes it last as the visitor leaves, while in the resident-resident case one is being served first and then the second resident is followed by an empty trial). Thus, the extended-credit model would assign somewhat higher value to the visitor and, subsequently, can develop some preference towards the visitor according to its decision rule.

759

760
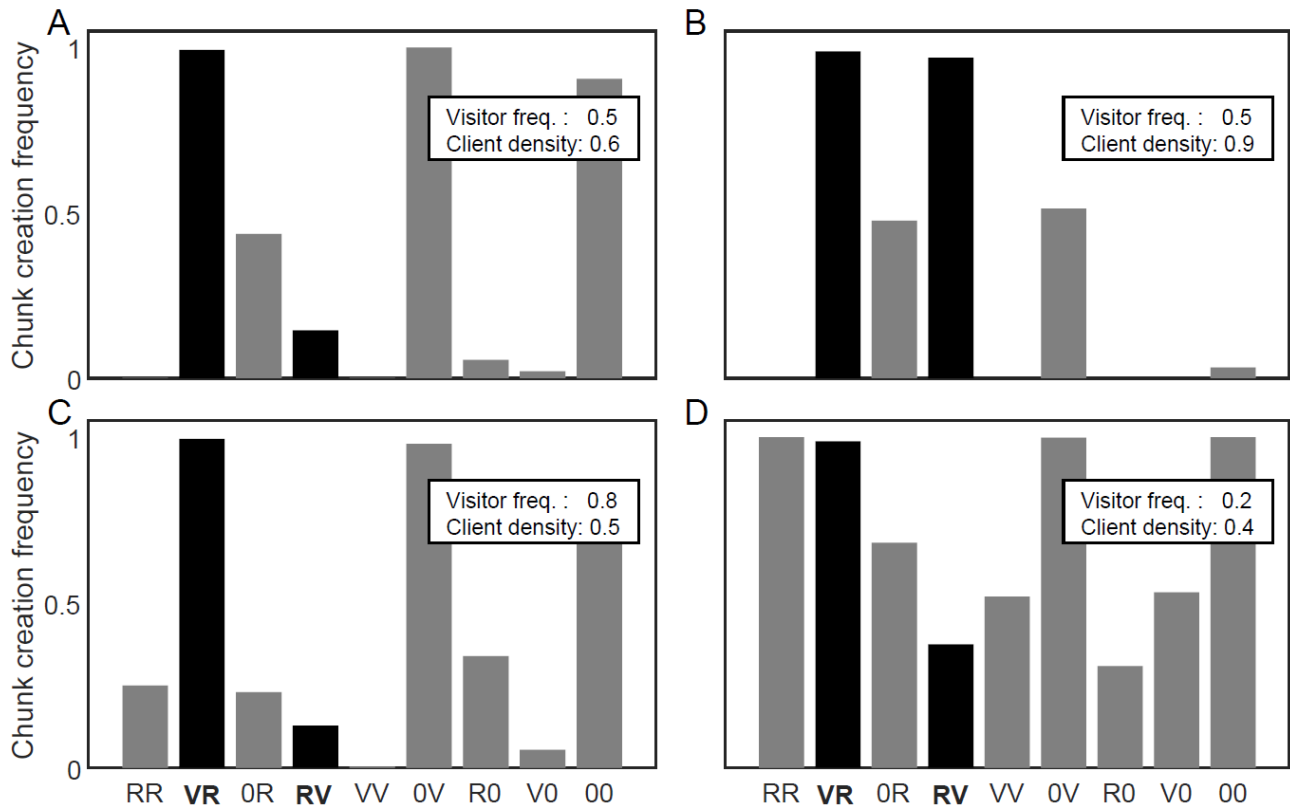
Figure S1. Creation of different chunks as part of the internal representation of the model in different ecological conditions. Four examples of ecological conditions are presented: **A)** visitor frequency of 0.5 and client density of 0.6, **B)** visitor frequency of 0.5 and client density of 0.9, **C)** visitor frequency of 0.8 and client density of 0.5, and **D)** visitor frequency of 0.2 and client density of 0.4. 1000 simulations were executed using the optimal C_p value for each condition (see Fig. 5A). The frequency of simulations, out of all simulations, in which the chunk was created by the end of the simulation, is presented for each chunk. Black bars – chunks which are relevant for the decision process; grey bars – chunks which are irrelevant for the decision.

761

762
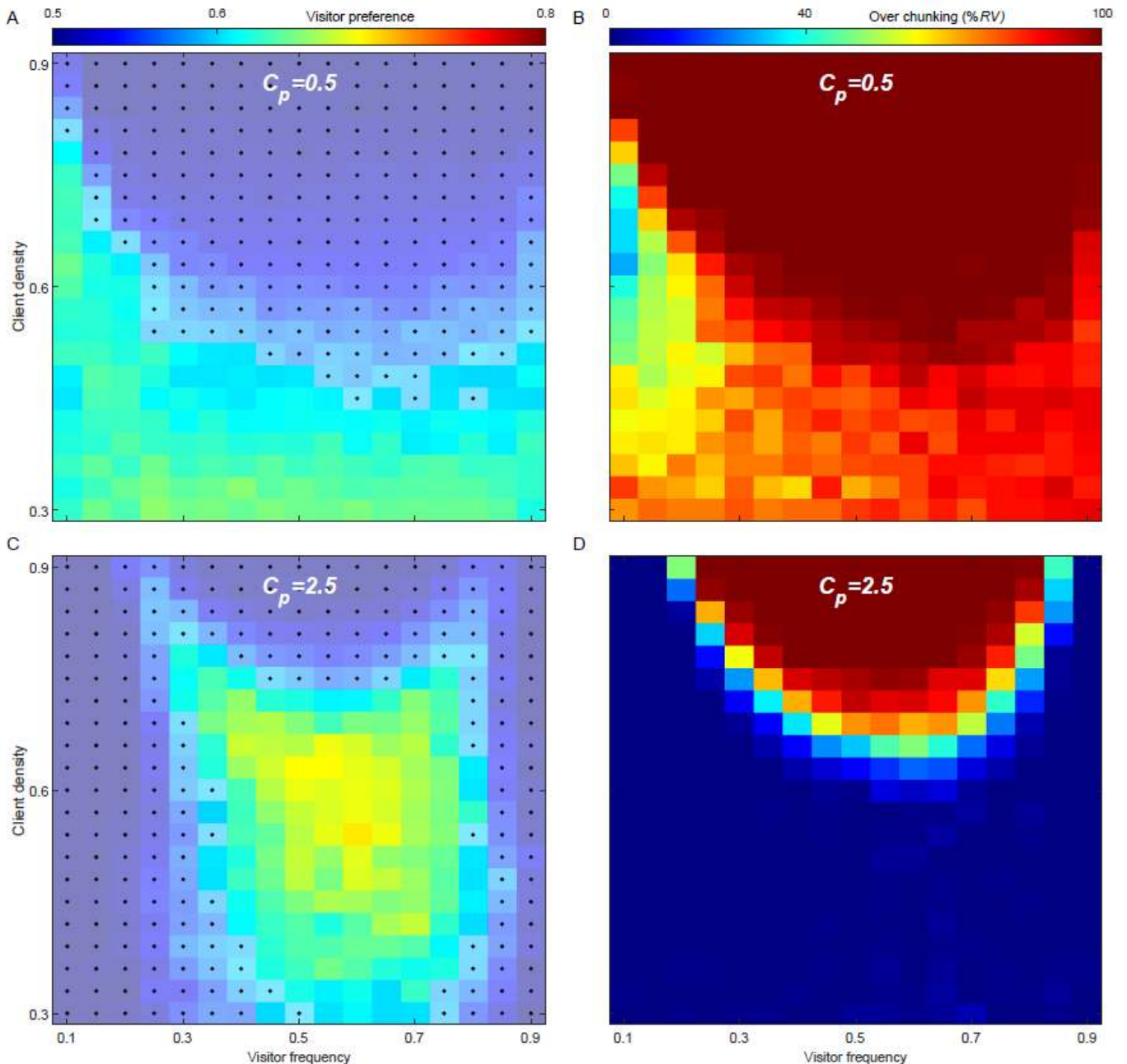
Figure S2. The link between ecological conditions, the success of the chunking model in the *natural market problem* using high and low C_p values, and over-chunking. **A)** The preference (colour) towards the visitor client when the $C_p = 0.5$ (low value) is used in different ecological conditions: the visitor frequency, $\frac{P_V}{P_V + P_R}$ (the fraction of visitors out of all clients), and the overall client density, $1 - P_0$. The preference at each point is the mean of 100 simulations. Light colors with black dots depict conditions in which the preference towards the visitor client is less than 0.6. **B)** The percentage of simulations which ended up with the model generating the maladaptive *RV* chunk (over-chunking), when $C_p = 0.5$. Computed using 100 simulations for each point. **C)** The preference towards the visitor client when the $C_p = 2.5$ (high value) is used in different ecological conditions. **D)** The percentage of simulations which ended up with the model generating the maladaptive *RV* chunk (over-chunking), when $C_p = 2.5$. Note that low C_p and high C_p are beneficial under different conditions. Over-chunking is the cause of failure in the low C_p case. On the other hand, in the high C_p case, over-chunking is responsible to failures only in some conditions (high client density), but under-chunking fails the model in other conditions (low visitor frequency).

763

764

765
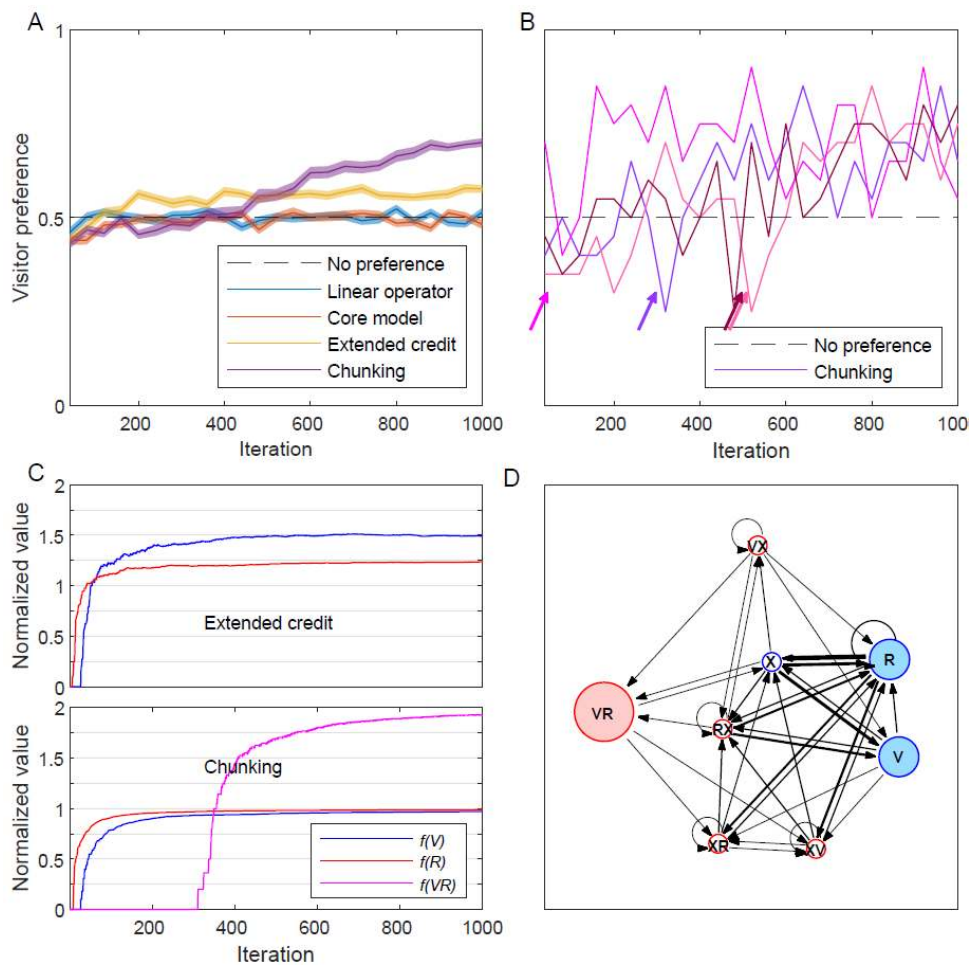
Figure S3. Simulating the *lab complex market problem*. A) Four types of learners are compared in the *lab complex market problem*: blue – A linear operator learner ($\alpha=0.1$; see text); orange – the core model; yellow – the extended credit model; purple – the chunking model (with $C_p=2$); black dashed-line – the expected choices with no preference (0.5). The preference towards a visitor client, measured as the proportion of choosing a visitor out of all visitor-resident encounters, is plotted as a function of time (iterations), in bins of 40 trials. Simulations are of 1000 feeding trials (with an empty trial after each feeding trial). The plots depict the mean of 100 simulations for each learner (shades – standard error of the mean) **B)** Four simulations of the chunking model. Note how the preference towards a visitor sharply increases after the creation of the *VR* chunk (depicted with an arrow for each simulation). **C)** The value of the different cues as perceived by the extended credit model (top) and the chunking model (bottom): blue – *V*; red – *R*; magenta – *VR*; in a single simulation. Note, how the extended credit model (top) converges towards a value of ~1.5 for *V* and 1.25 for *R*, giving rise to a slight preference (~0.6) towards a visitor (indicated by the yellow line in **A**; see text for discussion). **D)** The internal representation of the chunking model at the end of the simulation presented in (**C**, bottom). Blue – basic (initial) elements, red – chunk elements, filled nodes – the relevant elements for the decision process. The size of the circle is relative to the value (association with food reward) of the element. The width of the directed edges (black arrows) represents the relative frequency of the transitions between states (normalized *W*).

766

767

768   **Supplementary file:**

769   SimuFish.m – A Matlab function for running a simulation of the model in the cleaner fish market
770   problem. See documentation inside.

771