

## RESEARCH ARTICLE

# Modelling how cleaner fish approach an ephemeral reward task demonstrates a role for ecologically tuned chunking in the evolution of advanced cognition

Yosef Prat<sup>1\*</sup>, Redouan Bshary<sup>1</sup>, Arnon Lotem<sup>2\*</sup>

**1** Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland, **2** School of Zoology, Faculty of Life Sciences and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

\* [yosefprat@gmail.com](mailto:yosefprat@gmail.com) (YP); [lotem@tauex.tau.ac.il](mailto:lotem@tauex.tau.ac.il) (AL)

## OPEN ACCESS

**Citation:** Prat Y, Bshary R, Lotem A (2022) Modelling how cleaner fish approach an ephemeral reward task demonstrates a role for ecologically tuned chunking in the evolution of advanced cognition. *PLoS Biol* 20(1): e3001519. <https://doi.org/10.1371/journal.pbio.3001519>

**Academic Editor:** Simon W. Townsend, University of Zürich, SWITZERLAND

**Received:** June 22, 2021

**Accepted:** December 20, 2021

**Published:** January 5, 2022

**Copyright:** © 2022 Prat et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its [Supporting Information](#) files.

**Funding:** This study was supported by the Swiss National Science Foundation (grant number 310030B\_173334/1 to R.B.), and Israel Science Foundation (grant number 1126/19 to A.L.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## 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 mechanism enabling chunking ability and the conditions for its evolution and success, based on the ecology of the cleaner fish *Labroides dimidiatus*. In some environments, 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 this task in its common simplified laboratory format 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.

## 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” (a term that is commonly used but is loosely defined [1]) and exploring the ecological conditions that could possibly favour their evolution (e.g., [2–7]). Accurate navigation [8], social manipulations [9], or flexible communication [10], for example, may all be considered advanced cognitive abilities. Yet, mapping these skills along phylogenetic trees and their relation to social or ecological conditions (e.g., [11,12]) does not explain how such abilities evolved through incremental

**Competing interests:** The authors have declared that no competing interests exist.

modifications of their mechanistic building blocks. Earlier views of cognitive evolution were based on some postulated, loosely defined genetic adaptations, such as language instinct [13,14], mind reading abilities [15], or mirror neurons [16,17], but those are increasingly replaced by approaches relying on explicit associative learning principles that can gradually form complex representations of statistically learned information [18–26]. 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, so that they can improve decision-making and eventually enhance fitness. In other words, understanding the evolution of cognition requires to explain cognitive abilities first in terms of their possible mechanisms (proximate level of explanation) and then in terms of how such mechanisms could have evolved as a result of gradual modifications that improve biological fitness (ultimate level of explanation).

A relatively simple and well-understood example is the extension of simple conditioning through second-order conditioning in a process known as chaining [27,28]. 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 [29,30] or even social learning [31], which can clearly be adaptive.

Further elaborations of associative learning that may allow to construct a detailed representation of the environment and to support statistical learning and decision-making, such as those required for learning visual or vocal patterns [32,33], grammatical rules [6,34], or for the planning of sequential actions [35–37] are less well understood. It has become clear, however, that a critical requisite for such cognitive skills is the ability to represent 2 or more data units as a group that has a meaning that is different from (or independent of) the meaning of its components (as in the word ‘carpet’, which is not related to ‘car’ or ‘pet’). This ability has appeared in the literature under different names, such as ‘configurational learning’ [38,39], ‘chunking’ [23,40,41], or ‘segmentation’ [42], all of which are quite similar and involve the learning of configurations, patterns, and hierarchical structures in time and space [43].

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 [44]). Configurational learning of this type is contrasted with ‘elemental learning’, which is based on the behaviour expected from simple associative learning [45,46]. Research on configurational learning has been focused mainly on identifying the brain regions supporting this ability (e.g., [39,47–49]), giving relatively little attention to the cognitive processes generating configural representations (but see [39]). More attempts to consider these possible processes has been made in the context of chunking or segmentation (e.g., [18,42,50]), but only recently, theoretical work has started to address the question of how chunking mechanisms evolve under different ecological conditions, and what is their role in cognitive evolution [23,51,52].

A unique model system that may provide a remarkable opportunity to study the evolution of chunking is that of the bluestreak cleaner wrasse (*Labroides dimidiatus*), which feeds on ectoparasites removed from “client” fish [53]. Field observations and laboratory experiments have shown that at least some of these cleaner fish are capable of solving a problem known as the market problem (or the ephemeral reward task) [54–57]. The market problem entails that if approached by 2 clients, cleaners must learn to serve a visitor client before a resident client, because the latter waits for service while the former leaves if not attended. In other words, a preference for a visitor when approached by a visitor and a resident provides the cleaner with 2 meals while failing to do so may result in losing one of them (see details in [54,56,58,59]). In

order to choose correctly, the cleaner also has to distinguish between residents and visitors based on their appearance or behaviour (and there might be multiple client species acting as residents and visitors within a cleaner's home range [54]). In the lab, clients of different types were replaced with plates of different colours, each offering 1 food item and acting either a visitor or a resident, which was sufficient for some of the cleaners to solve the problem correctly [55,60]. These experiments suggest that cleaners can distinguish visitors from residents by associating certain visual cues with their previous behaviour. Interestingly, individuals captured in different habitats demonstrated different learning abilities of the market problem in the lab, and adult cleaners seem to learn better than juveniles [56,58,61,62]. Such intraspecific variation in cognitive abilities suggests some role for the ecological and the developmental circumstances in the fish life history.

The lab market task may first appear as a two-choice experiment, testing whether animals can learn to choose the option that yields the largest total amount of food. Nevertheless, while preferring a larger amount in a simple two-choice task seems almost trivial for most animals [63], the market version, in which a double amount is a product of a sequence of 2 actions (i.e., choosing ephemeral and then approaching the enduring item) has been proven difficult for a range of species [62,64–66] (but see [67–69]). Follow-up studies on pigeons and rats (reviewed in [70]) showed that letting the subject make a first decision but delaying the consequences, i.e., delaying the access to the rewarding stimuli, strongly improves performance [70,71]. One interpretation of these results is that the delay helped animals to connect their initial choice to both consequences; the first, and then the second reward, both of which occurred within a short time span after the relatively long delay.

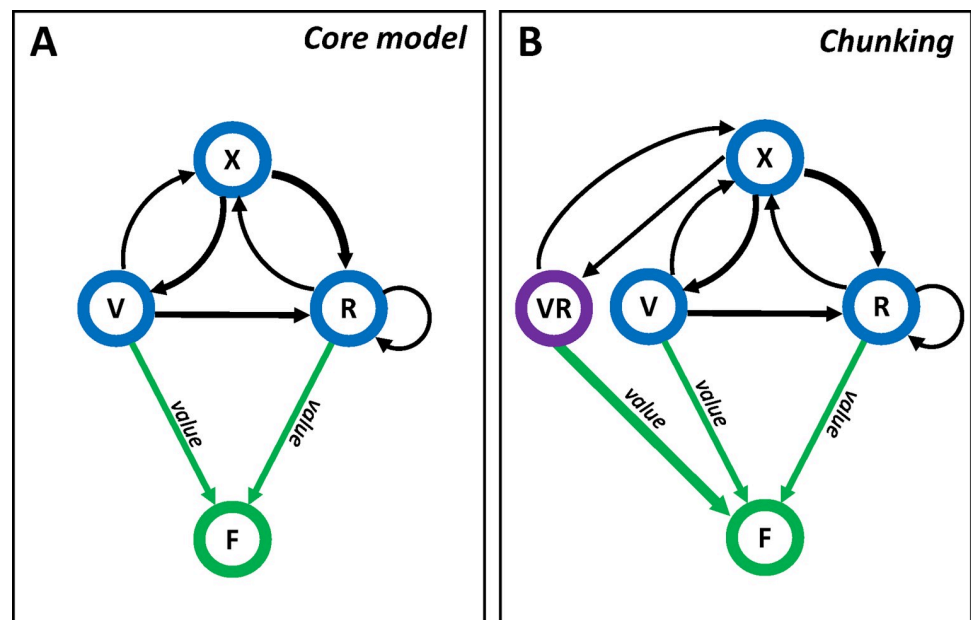
While delaying the consequences of the initial choice may be helpful under some conditions, recent theoretical work suggests that under natural conditions, basic associative learning is insufficient for solving the market task, which instead warrants some form of chunking ability [72]. The reason for that is that the commonly used laboratory task presents a relatively simple version of the problem compared to the natural situation. It only presents visitor–resident pairs, for which choosing the visitor first always entails double rewards and choosing the resident first always entails a single reward. In nature, however, cleaners face also resident–resident as well as visitor–visitor pairs, and most often, only a single client approaches. As a result, choosing visitor first may not always entail double reward (e.g., in visitor–visitor pairs, the second visitor is likely to leave) and choosing resident first may not always result in a single reward (e.g., in resident–resident pairs, the second resident is likely to stay). Indeed, the theoretical analysis carried out by Quiñones and colleagues [72] showed that for solving the natural market problem, it is necessary to have distinct representations of all different types of client combinations (visitor ( $v$ ) + resident ( $r$ ),  $r + r$ ,  $v + v$ ,  $r$ , and  $v$ ), which means the ability to represent chunks. Yet, the analyses did not explain how such representations are created, and to what extent ecology causes variation in the cleaners' ability to create such representations. The goal of the present paper is to fill up this critical gap.

Following Quiñones and colleagues' demonstration that chunking is necessary for solving the natural market problem, here we use the cleaner fish example as a means to study the evolution of an explicit chunking mechanism and the ecological conditions that favour its success. Thus, we investigate how the very same problem—choosing between 2 options where one yields the double amount of food—set into an increasingly complex ecology (of facing varying combinations of these options) selects for the evolution of increasingly advanced associative learning abilities. Our model is based on a weighted directed graph of nodes and edges, which initially form a simple associative learning model, and can then be modified to become an extended credit (chaining-like) model, or a chunking model (see details and definitions below). This approach allows to compare between clearly defined learning mechanisms and to

pinpoint the modifications responsible for a presumed evolutionary step that improves cognitive ability. We analyse the 3 learning models' performance in 3 tasks: the basic quantitative choice task, the laboratory market task, and the market task embedded in a sequence of varying configurations ("natural market task"). For the latter, we explored to what extent different densities and frequencies of client types select for different tendencies to form chunks (a critical parameter in the model), and how such different tendencies may affect the cleaners' ability to solve the market problem.

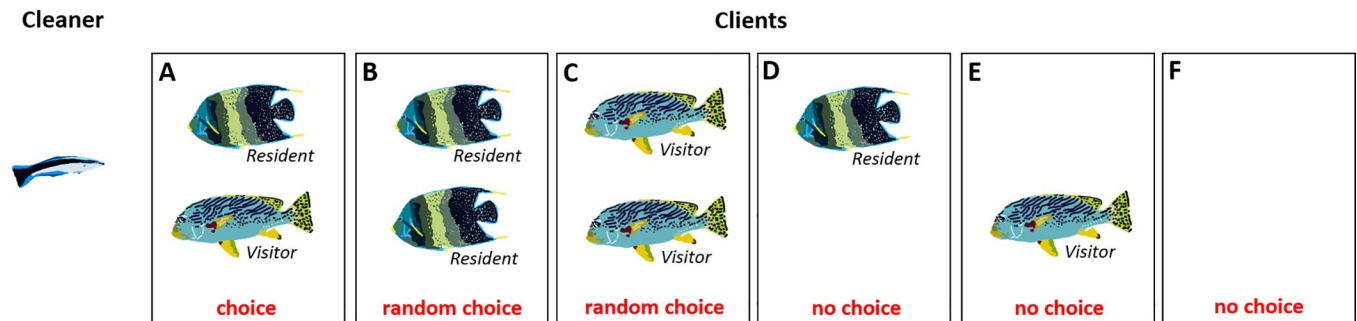
## The core model

**Internal representation.** Our core model consists of a weighted directed graph  $G = (N, E)$ , with nodes  $N$ , edges  $E$ , and additionally edge weights  $W$ , node weights  $U$ , and node values  $F$  (Fig 1A). The basic model includes 3 internal nodes representing 3 behavioural states:  $N = \{V, R, X\}$ , where:  $V$ —serving (feeding on) a visitor—client,  $R$ —serving a resident—client, and  $X$ —waiting for clients (empty arena). These are the 3 states (responses to environmental cues) required to represent the market problem and are therefore available to, and perceived by, the cleaner fish in our simulations (Fig 2). Note that at this stage, the cleaner does not understand the behavioural differences between a resident and a visitor, yet we assume it can distinguish between their external characteristics (e.g., the cleaner can identify their colours as different colours). Edge weights are updated according to the sequential appearance of the states, i.e., whenever  $n_j$  appears after  $n_i$  the weight of the edge  $n_i \rightarrow n_j$ , i.e.,  $W(n_i, n_j)$ , increases (by 1 unit, in our simulations). Thus, edge weights represent the associative strength between nodes experienced one after the other. For simplicity, we ignore weight decay (forgetting) in the present model (see Discussion section where we address this issue and explain why it should not



**Fig 1. Model design—internal representation.** (A) The core model contains a network of 3 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  $R$ ”.

<https://doi.org/10.1371/journal.pbio.3001519.g001>



**Fig 2. Model simulations.** The cleaner in our simulations may encounter different combinations of client pairs awaiting its service: (A) the cleaner must choose between 2 clients of different types according to the model's decision process; (B and C) the cleaner chooses with equal probabilities between 2 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.

<https://doi.org/10.1371/journal.pbio.3001519.g002>

change significantly the results). Node weights and values are attached to the cleaner's decisions (see decision-making below) according to their occurrence and association of their outcome with food, i.e., whenever node  $n_i$  is chosen, the weight  $U(n_i)$  increases (by 1 unit, in our simulations), and the value  $F(n_i)$  increases by the amount of food reward provided (which, unless otherwise specified, is assumed to be 1 unit per client if served successfully, and zero otherwise). The value of a node could be regarded as the strength of its association with food, which can also be represented as the weight of the edge between the node and a reinforcer food node (the weights of green arrows in Fig 1A). The weighted directed graph constitutes the cleaner's internal representation of the market environment. The cleaner's decisions regarding which clients to serve depend only upon this representation (Fig 1A).

Initially, the states are considered unknown to the fish and their corresponding values, weights, and the weights of their connecting edges are set to zero. Most learning models use prior values for cues or states, which are commonly set to zero (often implicitly). Here, we model such a prior by imposing a threshold on the weight of a node before any increase in its value  $F$  can occur. Specifically,  $F(n_k)$  is initialized to zero and would not change as long as  $U(n_k) < Q$ , i.e., at the first  $Q$  occurrences of  $n_k$ . We set  $Q = 10$  throughout all simulations, which implies that the value of a node will increase above zero only from the 11th serving of a client.

**Decision-making.** When a cleaner fish is presented with 2 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 2 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  $F(V)$ . As described by Eq 1, a soft-max function is employed (see [72]) over the 'normalised values'  $f(n_i)$  and  $f(n_j)$  such that the probability of choosing  $n_i$  is:

$$\pi_i = \frac{1}{1 + e^{-(f(n_i) - f(n_j))}} \quad (\text{Eq1})$$

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 state  $n_k$  (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 [73], which is a basic and widely used learning model [74] that does not involve chaining or chunking. The learner updates the value  $f(i)$  of state  $i$  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 state  $i$  at time  $t$  and  $\alpha$  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).

### The extended credit model

A straightforward approach to consider higher-order dependencies is to enable association of states with their “future” rewards. We call this model the “extended credit” model. The network representation of the extended credit model is the same as that of the core model ([Fig 1A](#)), but in this model, the learner associates an obtained reward with the current state as well as with the previous one. Specifically, while 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  $F(n_i)$  and  $F(n_j)$  increase (i.e., the credit assignment of the reward is extended also to the previous state). Hence, if both states are similarly rewarding, the first one will be associated with double the food by the end of the sequence, as it was also associated with a delayed reward. Theoretically, credit assignment could be extended in more than one step backward and the credit could also change (e.g., decrease) with time (similarly to “chaining” [75]). Note that although the model extends the credit to a previous state, it does not represent, in the credit extension, the **identity** of the consecutive state, which donated the extra reward. Thus, the extended credit model cannot learn to distinguish between different sequences (sequential combinations or configurations) of states (e.g.,  $V \rightarrow R$ ,  $V \rightarrow X$ ,  $R \rightarrow R$ , etc.). The decision-making process of the extended credit model is the same as in the core model (see above).

### The chunking model

Another way of identifying high-order dependencies is via configurational learning, or chunking, as mentioned in the Introduction section. To model how acquired experience leads individuals to create chunks, we employ a chunking procedure in our model in which sequences occurring more often than expected, according to the distribution of their elements are “chunked” into a new element ([Fig 1B](#)). Specifically, a sequence  $(n_i, n_j)$  would become a new element “ $n_i n_j$ ” of the internally represented network (i.e., a new node in the graph  $G$ ) whenever

$$W(n_i, n_j) > M \cdot P(n_i) \cdot P(n_j) + C_p \cdot \hat{\sigma} \quad (\text{Eq2})$$

where  $W(n_i, n_j)$  is the number of observed occurrences of the sequence  $n_i \rightarrow n_j$ ,  $M$  is the total number of observed states (or pair sequences),  $P(n_k)$  is the observed frequency of the element  $n_k$ , and  $\hat{\sigma}$  is the standard deviation of a binomial distribution, with the probability of an event  $n_i \rightarrow n_j$  being  $P(n_i)P(n_j)$ :

$$\hat{\sigma} = \sqrt{M \cdot P(n_i)P(n_j) \cdot (1 - P(n_i)P(n_j))} \quad (\text{Eq3})$$

$C_p \geq 0$  is a chunking avoidance parameter. This parameter is important, as it governs the behaviour of our model, or in other words, the conditions under which a chunk will be created (see [Discussion](#) section for possible implementations of this parameter in the brain). Note that when  $C_p = 0$ , any slight above chance co-occurrence of  $n_i$  and  $n_j$  would result in chunking. This is probably too much chunking because it can easily happen in nature for almost any 2 elements as a result of stochastic deviations from the frequency expected by chance. Using a  $C_p$  that is greater than zero implies that a chunk will be created only when the co-occurrence is higher than expected by a certain threshold.

Additionally, chunks would not be created as long as  $W(n_i, n_j) < Q$ , i.e., during the first  $Q$  occurrences of the sequence  $n_i \rightarrow n_j$ . This rule enforces a minimal sample size before statistical inference could be done (for chunking).

In this model (see [Fig 1B](#)), whenever a chunk is created, it is treated as a new node and is being associated with food whenever chosen by the cleaner alongside food reward (but only after its first  $Q$  occurrences, as required for other elements). For instance, if the sequence  $V \rightarrow R$  is chunked into a new element “VR,” further choices of the sequence  $V \rightarrow R$  will increase the association of the element “VR” with the reward by 2 units (as this is the observed reward during the processing of the sequence). On the other hand, if the sequence  $R \rightarrow V$  is chunked into a new element “RV” (which could happen in the ‘natural market problem’; see simulated environments below), further choices of the sequence  $R \rightarrow V$  will usually increase the association of the element “RV” with the reward by 1 unit only (since the visitor leaves if not served first).

The decision-making process of the chunking model is the same as in the core model (see above), but here, more choices may become available. For example, after the chunk “VR” is created, a cleaner faced with a visitor and a resident client simultaneously can choose to serve the resident ( $R$ ) or to perceive them as the chunk “VR” and to execute the sequence  $V \rightarrow R$  (i.e., approach the visitor and then the resident). On the other hand, if the chunk “RV” was also created, an additional option exists, which is the choice of executing the sequence  $R \rightarrow V$ . Importantly, in this case, soon after approaching the resident, the visitor would leave the arena so the outcome of choosing and attempting to execute the sequence  $R \rightarrow V$  may end up with serving only  $R$  (depending on the simulated environment; see below) and being reward by only 1 unit (see above). We assume that if a chunk has already been created the cleaner never chooses the first element alone if presented with both elements (i.e., if “VR” is already represented in the network, and “RV” is not, the cleaner should only choose between “R” and “VR” when presented with both client types simultaneously).

## Simulated environments

Our simulations provide the cleaner fish with alternating clients awaiting its service ([Fig 2](#)). The simulated arena includes 2 available spots, where each can be occupied by a visitor client, a resident client, or remain empty (simultaneous encounters with more than 2 clients are relatively rare in nature and are typically not addressed; see [[54,55](#)]). Each simulation consists of a sequence of discrete trials. On each trial, the arena is filled using a random sample according to the simulation specific setup (i.e., the probabilities of encountering each client type, as will be explained below). When the cleaner is presented with an empty arena (none of the 2 spots is occupied), it perceives the corresponding state  $X$  (see [Fig 1A](#)) and waits for the next trial (the next occupancy of the 2 spots). When the cleaner is presented with only a single client, it immediately serves it, perceives the corresponding state (i.e.,  $V$  for choosing to serve a visitor, or  $R$  for choosing to serve a resident), experiences the associated reward of serving it, and waits for the next trial. When the cleaner is presented with 2 clients, it chooses one according

to the decision rule of the model (see above) if the clients are of different types (a visitor and a resident), or at random (with equal probabilities) if they are of the same type. The cleaner then serves the chosen one and perceives the corresponding state (i.e.,  $V$  or  $R$ ) and its associated reward. If the second client (not chosen) is a visitor, it leaves and the cleaner waits for the next trial, but if the second client is a resident, the cleaner serves it as well, perceives the corresponding state ( $R$ ) and its associated reward, and waits for the next trial. Recall that whenever the cleaner chooses to serve a client and perceives its associated food reward, it adds 1 unit to the value  $F$  of the corresponding state (to  $F(V)$ ,  $F(R)$ ,  $F(VR)$ , etc.).

We have simulated 3 different environments: (i) A laboratory environment with a “basic two-choice task,” where a cleaner has to choose between 2 clients offering a reward of 1 and 2 units, respectively, and no further approach to clients is allowed after this initial choice within a trial (only in this simulation, both client types are ephemeral). This two-choice task is expected to be solved by all types of cleaners (i.e., preferring the client offering the double amount of food), thus serving as a ground-level associative learning test. (ii) A laboratory environment with a “laboratory market problem,” where the cleaner faces a resident and a visitor client together (Fig 2A), in each feeding trial, and after it finishes feeding it faces a single trial of empty arena (Fig 2F; i.e., perceives the state  $X$ ). (iii) A natural setting, henceforth termed “the natural market problem,” in which the cleaner may face all possible combinations (Fig 2A–2F): a visitor and a resident, 2 residents, 2 visitors, a single client (resident or visitor), and no clients. In addition, in the natural setting, the cleaner does not necessarily have to wait between trials. This environment simulates more faithfully the situation in the wild, where each of the 2 spots is filled using an independent random sample, with a probability  $P_V$  for a visitor, a probability  $P_R$  for a resident, and a probability  $P_0$  for an empty spot ( $P_V + P_R + P_0 = 1$ ). When examining ‘the natural market problem’, we consider the distribution of the different client types and their combinations as resulting from 2 ecological parameters: the (relative) visitor frequency,  $\frac{P_V}{P_V + P_R}$  (the fraction of visitors out of all clients), and the overall client density,  $1 - P_0$  (ranging from zero—when there are no clients and the cleaner always faces an empty arena—to one, the arena is always full).

Simulations were performed using Matlab 9 (the code is provided in [S1 File](#)).

## Results

We examined how the 4 learning models fare in the 3 simulated tasks: the basic two-choice task, the laboratory market problem, and the natural market problem.

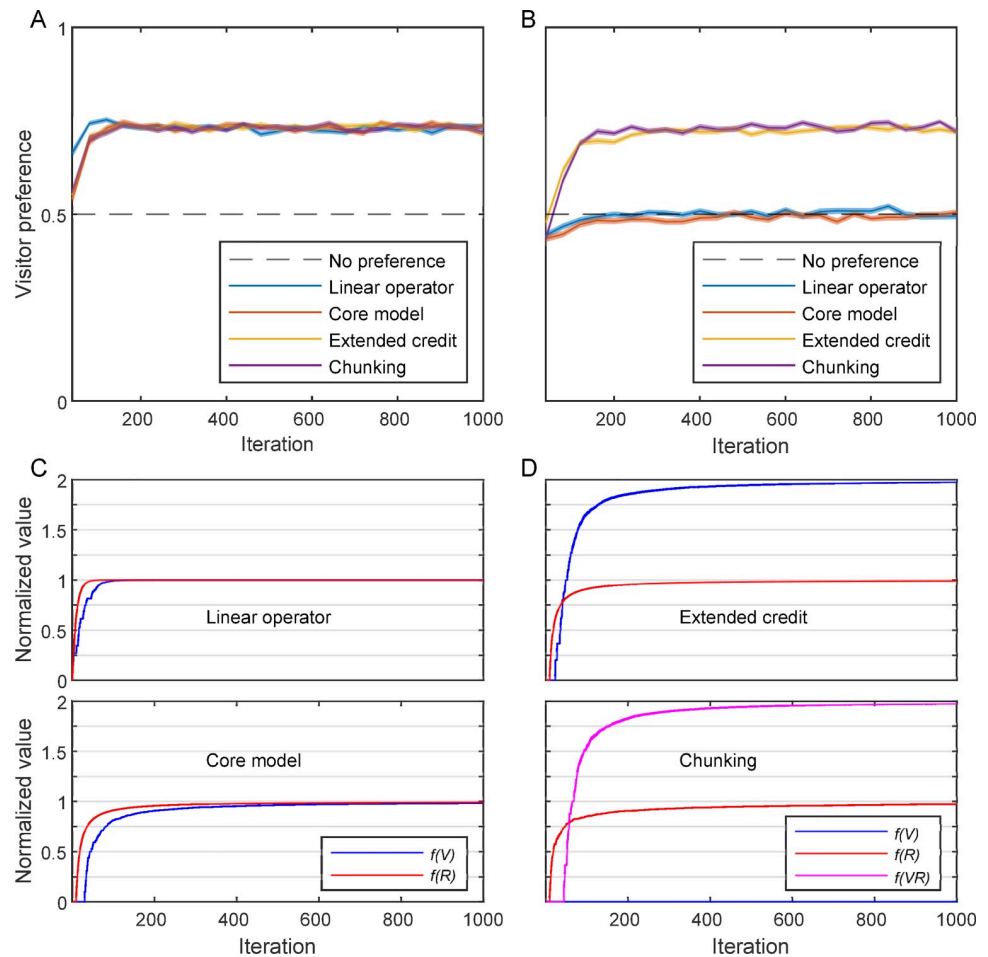
### The basic two-choice task

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

### The laboratory market problem

Facing the laboratory market problem, the core model and the equivalent linear operator learner did not develop a preference towards any of the given clients and thus failed to solve the problem (Fig 3B, orange and blue lines, respectively). In contrast, both the extended credit





**Fig 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 1,000 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 states as perceived by the nonsuccessful 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. (Underlying data in [S1 Data](#)).

<https://doi.org/10.1371/journal.pbio.3001519.g003>

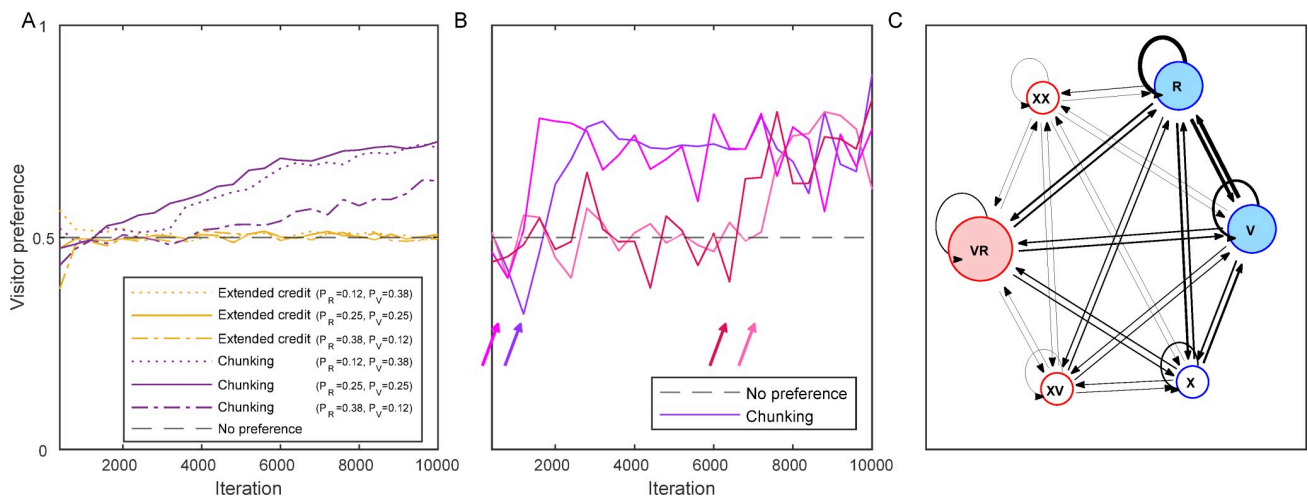
model and the chunking model were capable of solving the ‘laboratory market problem’, i.e., to develop a strong preference towards the visitor client (Fig 3B, yellow and purple lines, respectively). The inability of the core and linear operator models to solve the laboratory market problem is reflected in their indifferent ‘normalised values’ of R,  $f(R)$ , and V,  $f(V)$ , both of which approach 1 (Fig 3C). This result is expected since the value of each state is updated independently of any past and future state or reward, and both states (client types) provide the exact same immediate reward. In the extended credit model that solves the problem successfully, the ‘normalised value’ of R,  $f(R)$ , approaches 1, while the ‘normalised value’ of V,  $f(V)$ , approaches 2 (Fig 3D, top panel). This was made possible because serving a resident always provides a single food item in this setup (as the visitor leaves) while the credit of choosing a

visitor is extended to the resident that waits to be served (thus crediting  $V$  with 2 food items). The success of the chunking model is based on a different process: It creates the chunk  $VR$  early and  $f(VR)$  quickly approaches 2 as the complete chunk provides 2 food items (Fig 3D, bottom panel), pushing the preference towards a visitor client (the model chooses the sequence  $VR$ , i.e.,  $V$  and then  $R$ ).

## The natural market problem

For the natural market problem, we only present the results of the learning models that were successful in solving the laboratory market problem (as expected, the core and the linear operator models that failed to solve the laboratory problem also fail to solve the more complex natural problem).

The extended credit model that was sufficient for solving the laboratory market problem failed to solve the natural market problem (i.e., to prefer a visitor client) regardless of the overall client density or the relative frequencies of different clients (see examples in Fig 4A, yellow lines). The reason for that is that in the ‘natural market problem’ all pair sequences can occasionally appear, including a resident after a resident (and thus  $R$  is credited with 2 food items), a visitor after a visitor (and thus  $V$  is credited with 1), and even a resident and then a visitor (when there is no empty trial after serving the resident, which again credit  $R$  with 2). Thus, assigning credit for a state for the value of the next state causes the differences between  $f(R)$  and  $f(V)$  to vanish. Still, the sequence  $V \rightarrow R$  may occur more often than the sequence  $R \rightarrow V$  (at least as long as the cleaner do not prefer  $R$ ), since whenever the 2 types of clients appear simultaneously,  $V \rightarrow R$  occurs if the cleaner chooses to serve the visitor first, and  $R \rightarrow V$  occurs only when a visitor appears by chance in a new trial after the cleaner has served a resident. As a result,  $f(V)$  might be slightly greater than  $f(R)$  in some situations. However, in order to respond to such slight differences, the model’s soft-max decision rule should be ‘hardened’ (become more similar to a maximum-based rule). This would suppress exploration and make the



**Fig 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 ( $I-P_0$ ) of 0.5 and for 3 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), and  $P_R = 0.38$  and  $P_V = 0.12$  (dotted dashed lines).  $C_p = 2$  (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. (Underlying data in S1 Data).

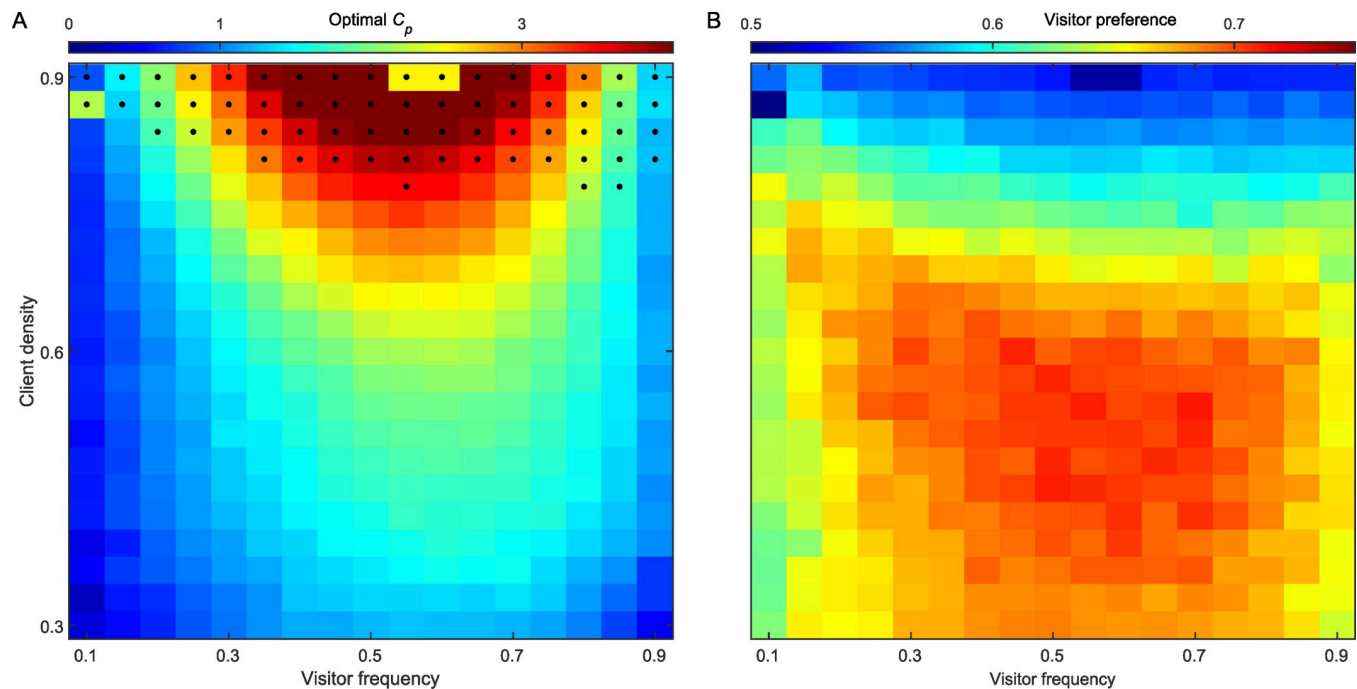
<https://doi.org/10.1371/journal.pbio.3001519.g004>

model always choose the most frequent client type (as its value increases faster), which is the resident in most cases since the visitor leaves if not served.

In contrast to the extended credit model, the chunking model solved the ‘natural market problem’ successfully in a wide range of client distributions (Fig 4A, purple lines). To solve this task, the chunking model only needs to create the chunk  $VR$ , which, in turn, imposes a preference for the visitor, as  $VR$  is always associated with 2 units of food reward. The time of creating the  $VR$  chunk may vary according to the stochastic order of the trials experienced by each individual (see examples in Fig 4B). But on average, as the simulation advances, the chances of a cleaner using the chunking model to create the  $VR$  chunk and thus to choose a visitor increases (Fig 4A, purple lines). Fig 4C depicts an example of the internal representation of the chunking model at the end of a simulation of the ‘natural market problem’. Note that the chunking model creates chunks regardless of the reward, and depending only on the statistics of state occurrence. Thus, it may also create chunks containing the state for an empty arena ( $X$ ), or for other various combinations ( $XR$ ,  $XV$ ,  $RR$ ,  $VV$ , etc.). In most cases, these chunks do not influence the cleaner’s decisions as they represent states that require no choice (see Fig 2). However, as we shall see below, in the natural setting, there is also a risk of creating the  $RV$  chunk (rather than  $VR$  chunk), which can bias the cleaner’s decision, implying that chunking should be limited to avoid overchunking.

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

The behaviour of the chunking model is controlled by the chunking avoidance parameter  $C_p$  (Eq 2). Large values of  $C_p$  prevent any chunking and the model is reduced back to the core model (which do not develop a preference towards the optimal choice). On the other side, too low values of  $C_p$  cause “overchunking.” Therefore, the optimal value of  $C_p$  will depend on the ecological conditions: the overall client density and the frequency of the different client types. If there are many clients per cleaner, cleaners will often be solicited. Therefore, a visitor may regularly appear right after a resident—not because the visitor waits for service, but because a new visitor client enters into the arena by chance. Thus, there is a risk that the misleading chunk  $RV$  might be created, as well as the beneficial chunk  $VR$ . The reason we view the  $RV$  chunk as misleading is that faced with a choice between a visitor and a resident, the cleaner can now consider both sequences of actions,  $V \rightarrow R$  and  $R \rightarrow V$ , and choose between them according to their expected values. Although the value of the chunk  $VR$ ,  $f(VR)$ , would approach 2 and hence be higher than the value of the chunk  $RV$  (with  $f(RV)$  lower than 2), the decision rule allows some proportion of choosing the  $RV$  chunk (exploration), which result in serving the resident first. In other words, overchunking reduces the strength of the preference for the optimal choice. The balance between underchunking and overchunking implies the existence of optimal  $C_p$  values (balancing between the need to create the  $VR$  chunk but not the  $RV$  chunk). Importantly, these optimal  $C_p$  values depend on 2 ecological conditions: the overall client density and the frequency of the different client types, which determine how frequently the sequences  $V \rightarrow R$  and  $R \rightarrow V$  are likely to be encountered. The effect of these ecological conditions on  $C_p$  and on the success of solving the ‘natural market problem’ is shown in Fig 5. It can be seen that in some extreme ecological conditions (of high client densities), it would be difficult for a cleaner fish using the chunking model to solve the market problem with any  $C_p$  (Fig 5A, black dots, and Fig 5B, blue shades representing low preference for visitors), since empty spots are rare events and most choices of the chunk  $RV$  result in obtaining 2 units of food (from the resident and the subsequent served client from the next trial). Fortunately for the cleaners, solving the market problem under these high client density conditions is not



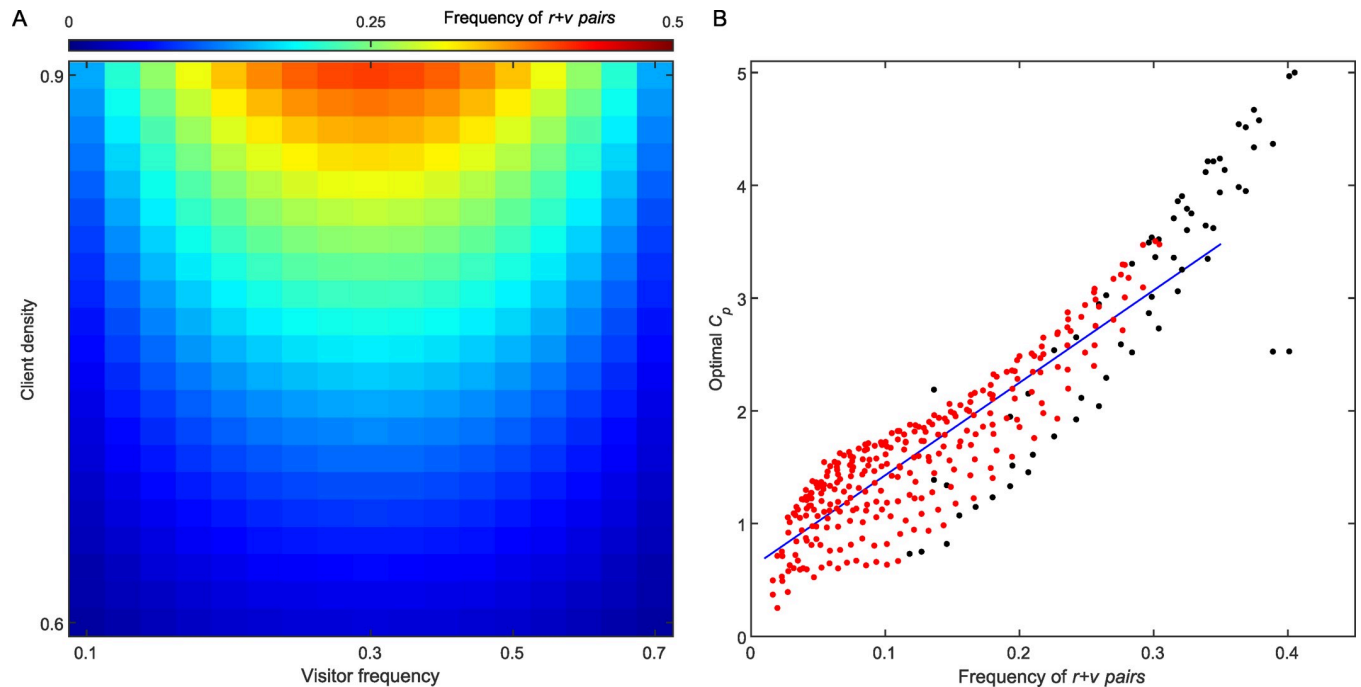
**Fig 5. The link between ecological conditions, optimal  $C_p$ , and the success of the chunking model in the ‘natural market problem’.** (A) Optimal  $C_p$  values (that provide the strongest preference towards a visitor), indicated by colour, as a function of 2 ecological conditions: the visitor frequency,  $\frac{p_V}{p_V+p_R}$  (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 1,000 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. (Underlying data in S1 Data).

<https://doi.org/10.1371/journal.pbio.3001519.g005>

important in nature as high client densities lead to near permanent demand for cleaning. Yet, in most simulated ecological conditions where solving the market problem is important, an optimal  $C_p$  value (Fig 5A) that induced a preference towards a visitor (Fig 5B) was found.

To visualise the importance of the overchunking problem, S1 Fig presents the frequency of appearances of each possible chunk (among 100 simulations) in 4 different ecological conditions, showing that when the chances of generating both the VR and RV chunks are similar (S1B Fig), the preference towards visitors vanishes (compare with the relevant point of 0.5 visitor frequency and 0.9 client density in Fig 5B).

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



**Fig 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;  $R^2 = 0.78$ ). (Underlying data in S1 Data).

<https://doi.org/10.1371/journal.pbio.3001519.g006>

## Discussion

Chunking mechanisms are essential to represent structured data in the brain and have probably played a pivotal role in the evolution of cognition [38,40,43,51,52]. Yet, a possible challenge in the evolution of chunking is that incorrect chunking and overchunking may lead to maladaptive behaviours and to cognitive impairments [76,77]. Indeed, the problem of under or overchunking arises whenever sensory input has to be chunked or segmented (reviewed in [43]). Normally, the problem is difficult to track because incoming data can be chunked in multiple ways and the number of possible chunks grows exponentially with the amount of data. This problem is well appreciated, for example, in the case of word segmentation during language learning in humans [42,78] or in the representation of behavioural sequences by animals [79]. Our analyses show that the market problem solved by cleaner fish in the wild offers a relatively simple model system to study the evolution of chunking. It is not only simple and tractable, but it involves a case where the function of chunking and its fitness consequences are well understood and are ecologically relevant, the adaptive and maladaptive chunks can be clearly identified (i.e.,  $VR$  versus  $RV$ ), and it can be studied experimentally and in relation to variable ecological conditions (e.g., [54,60,61]).

We implemented this approach by placing the same general problem of making a decision that doubles food intake in different sequential contexts that cleaners face in the wild. We show how solutions depend on increasingly complex learning rules. A simple two-choice task can be solved with basic reinforcement learning models such as the linear operator or our equivalent core model. A more challenging task where doubling the amount of food is consistently due to consequences of an initial choice (i.e., the ‘laboratory market task’) requires an

extended credit learning model that picks up a consistent chain of events. Finally, if cleaners face diverse sequences of events, as in the ‘natural market problem’, relevant causal chains of subunits that lead to doubling the food intake must be identified and chunked so that the animal can optimise food intake.

We also demonstrate that when facing diverse sequences of events, having the ability to chunk may not be sufficient. It is critical that the tendency to create chunks, captured by the chunking parameter  $C_p$ , be adjusted to ecological conditions. Moreover, our simulations also show that under some extreme conditions, even the optimal chunking parameter may not be sufficient for developing a preference for the ephemeral reward. In the cleaners’ market problem, it happens when the probability of encountering the sequences of the useful and misleading chunks,  $VR$  and  $RV$ , respectively, is so similar that no chunking parameter can allow the creation of  $VR$  while preventing the creation of  $RV$ . As mentioned earlier, in the case of the cleaner fish, this may not be a problem because it happens under conditions of high client densities where preferring the ephemeral reward (i.e., visitors) is not necessary. It is yet to be studied how common are such conditions in other problems animals face in nature, and to what extent using the right chunking parameter is sufficient for successfully balancing the trade-off between under and overchunking.

Demonstrating the trade-off between adaptive chunking and overchunking yields a new perspective on the cognitive basis of cleaner fish “cleverness” in their choices of clients. Solving the natural market problem does not represent an “all or none” cognitive ability but rather the ability to correctly adjust a more basic cognitive ability, which is the ability to create chunks. As it stands, many animals are capable of creating chunks and configurations in their memory representation (see [Introduction](#)), but only those applying the chunking parameters suitable to the required conditions will solve the natural market problem. The trade-off between chunking and overchunking may also explain why chunking (and configurational learning) takes time and may thus be viewed as difficult. Our model suggests that there is nothing really difficult in creating chunks quickly but that the process of chunking evolved to be slow in order to prevent overchunking. Note that the idea that learning may evolve to be slow as a result of a trade-off is not new. It is implied in the optimization of learning rate parameters to balance between exploration and exploitation in reinforcement learning models [[80,81](#)] and was also suggested as a way to minimise recognition errors [[82,83](#)].

## The mechanism of chunking

Our chunking model specifies the statistical conditions required for the formation of chunks and describes how chunks are represented in the network (Eqs [2](#) and [3](#); [Fig 1B](#)). Yet, it does not explain how chunks are actually created. In other words, it does not explain how it happens that under the conditions specified by Eqs [2](#) and [3](#), a chunk in the network suddenly appears. While the neuronal coding of such information is still poorly understood [[39](#)], a fairly explicit implementation of the process of chunk formation using neuronal-like processes may be possible (see also [[84](#)]). We can think of the required number of co-occurrences of  $V$  and  $R$  that is represented by the left side of [Eq 2](#) as the weight of their associative strength. Accordingly, a chunk representing the sequence  $VR$  is created when the weight of the edge leading from  $V$  to  $R$  passes a certain threshold. The formation of a chunk may be a result of another node in the network that receives signals from both neuronal units (or more precisely, from  $R$  soon after  $V$ ) and thus increases in weight and becomes the “chunk node” representing the repeated occurrences of the sequence  $VR$  (as in [Fig 1B](#)). The threshold weight required for the creation of a chunk can thus act as the chunking parameter  $C_p$  in our model and be optimised in line with [Eq 2](#).

In our model, which was kept as simple as possible, we assumed that weight increases by 1 unit per observation and does not decay over time. Realistically, however, different combinations of weight adjustment rates (increase and decrease) determine the timing of crossing the threshold for chunk formation. For example, slow increase in weight with a relatively fast decay require frequent co-occurrences in order to reach the threshold, creating a test for the chunk's statistical significance [23,43,51,85]. Thus, the chunking parameter in our model can be implemented by several mechanisms. We can, hence, view this parameter (or parameters) more generally as those effecting the tendency to form chunks (or the tendency to use configurational rather than elemental learning). In that sense, the value of these parameters could be a derivation of mechanistic elements such as the rates of weight increase and decrease (decay) as previously suggested [23,51,52]. Note, however, that although increasing the decay rate can minimise overchunking, it also, at the same time, acts against the creation of relatively rare but adaptive chunks, so the trade-off between adaptive chunking and overchunking still remains (see, e.g., [23,51,52,77], where both memory and forgetting were considered as the basis for chunking).

The optimization of the chunking parameters to ecological conditions may occur over generations through selection acting directly on parameter values, or instead (or in addition) cleaners may have evolved phenotypic plasticity with respect to the chunking parameter. For example, a systematic loosening of the chunking parameters (i.e., varying them more freely) when in poor conditions, and fastening the parameters (i.e., stop altering them) when in good conditions may bring the chunking parameters to get fixated around the values associated with best performance. Another possibility is that cases where a visitor is leaving without waiting are experienced by the cleaner as aversive (a loss of a meal) and the aversive saliency of such events has evolved to reduce the chunking threshold (which increases the likelihood of chunking when solving the market problem is indeed necessary).

### Implications of our results for the interpretation of empirical studies

A major insight from our model in comparison to Quiñones and colleagues [72] is that animals only need the ability to detect chains of events (rather than chunking) in order to solve the laboratory market problem. Accordingly, it is not at all clear that differences between species in performance in the laboratory market task are due to different chunking abilities or different values of chunking parameters. It is hence important to use a more complex design of the market task (which resemble the natural setting for which chunking is necessary) on species that have solved some form of the laboratory task, i.e., cleaner fish, African grey parrots, and capuchin monkeys [64,69]. Truskanov and colleagues [86] designed such a task, exposing cleaner fish to 50% of presentations of visitor and resident ( $r + v$ ) plates as well as to 25%  $r + r$  and 25%  $v + v$  presentations. While a few cleaners solved this task, overall performance tended to be lower than in the standard laboratory market task. Applying our learning models to this nonstandard (complex) market task showed that the extended credit model yields at best a slight preference for visitors, while the chunking model yields high performance (see [S1 Text](#) and [S3 Fig](#)). The study by Truskanov and colleagues thus yields experimental evidence that (some) cleaner fish can chunk. The task could also be adapted to test whether imposing “early commitment” that helped pigeons in solving the standard laboratory market problem [70] can also help to solve the natural problem, for which chunking ability is needed. Alternatively, “early commitment” can only help in extending the credit given to the initial choice (to the second reward as well as to the first one), which can solve the laboratory market problem but not the natural one.

Based on our model and simulations, there are currently multiple ways to explain the documented intraspecific variation in cleaner fish performance in both the standard and the complex laboratory market tasks [56,58,61,86]. First, variation in the laboratory market task may be related to whether individuals solve the problem by chunking or by chaining (extended credit) mechanisms and to individual variation in the fine-tuning of the parameters of each mechanism. Second, assuming that cleaners use chunking to solve the tasks, variation in their performance may be attributed to some limitations or time lags in optimising the chunking parameters to current conditions in the field or to the specific conditions in the lab. Such limitations and time lags are expected for both genetic and phenotypically plastic adjustments because in the cleaners' natural habitat, client densities and visitor frequencies can vary greatly across years and microhabitats [58,61], causing both inter and intraindividual variation within individual lifetimes.

Importantly, these interpretations make related assumptions amenable for future testing. For example, that fast-solving cleaners use chunking even in the laboratory market task even though chaining would suffice, and that cleaners apply their field experience and developed  $C_p$  value to the lab task. Some empirical results are already in line with the second assumption. First, the best predictor of high cleaner performance in the laboratory task is high cleaner fish density in the field [56], which in terms of our model implies low client density (per individual cleaner) and therefore low optimal  $C_p$  that promotes faster chunking (see Figs 5A and 6). Second, although it has been found that on the average individuals with relatively larger forebrains are more likely to be found in areas where they frequently face the market problem [87], on a local scale, individuals with relatively larger forebrains performed according to what appears to be the locally best strategy: to solve the task if living in a high-cleaner density area and to fail the task if living in a low-cleaner density area [88]. In terms of our model, such high and low cleaner densities correspond to relatively low and high client densities that favour low and high  $C_p$  values, respectively (see Fig 5A). Thus, bringing such cleaners to the lab implies that those who adaptively developed low  $C_p$  in their natural habitat are more likely to pass the test than those who developed high  $C_p$  (that was also adaptive to their natural habitat), which may explain Triki and colleagues' results [88]. It would certainly be interesting to explore the relationship between solving the ephemeral reward task and brain neuroanatomy also in other species. Yet, the fine-tuning of the chunking process demonstrated by our model suggests that experiments in each species should be calibrated to the natural frequency of stimuli in nature. Otherwise, a failure to solve the task may only indicate a mismatch between laboratory and natural conditions.

## Conclusions and implications for the study of advanced cognitive abilities

The cleaner fish ability to solve the market problem has presumably evolved on the background of its unique ecology and may be rightfully viewed as a surprisingly advanced cognitive ability for a (small brain) fish. However, by modelling the learning mechanisms required for this remarkable ability, we tried to put the cleaner fish story within the broader context of cognitive evolution, viewing it as a potential model for the evolution of chunking mechanisms. While the importance of chunking is usually considered within cognitive systems that are already highly advanced, the simple setting of the market problem allowed us to explicitly analyse the process of chunk formation, elucidating the trade-off between creating useful and misleading chunks, and demonstrating the importance of adjusting the chunking parameters to ecological conditions. We hope that the approach taken here could eventually be applied in the study of other cognitive abilities, identifying the learning mechanisms and the fine-tuning of



their parameters required for their success, and mapping them not only along phylogenetic trees but also along evolutionary axes of explicit incremental changes in learning and cognitive mechanisms.

## Supporting information

**S1 File. SimuFish.m—A Matlab function for running a simulation of the model in the cleaner fish market problem.** See documentation inside.

(M)

**S1 Data. Underlying data for Figs 3–6 and S1–S3.**

(XLSX)

**S1 Text. Details of the simulations of the laboratory complex market problem.**

(PDF)

**S1 Fig. 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. A total of 1,000 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 that are relevant for the decision process; grey bars—chunks that are irrelevant for the decision. (Underlying data in S1 Data).

(EPS)

**S2 Fig. The link between ecological conditions, the success of the chunking model in the ‘natural market problem’ using high and low  $C_p$  values, and overchunking.** (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 colours 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 (overchunking), 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 (overchunking), when  $C_p = 2.5$ . Note that low  $C_p$  and high  $C_p$  are beneficial under different conditions. Overchunking is the cause of failure in the low  $C_p$  case. On the other hand, in the high  $C_p$  case, overchunking is responsible to failures only in some conditions (high client density), but underchunking fails the model in other conditions (low visitor frequency). (Underlying data in S1 Data).

(EPS)

**S3 Fig. 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 1,000 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 states 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 approximately 1.5 for  $V$  and 1.25 for  $R$ , giving rise to a slight preference (approximately 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 (normalised  $W$ ). (Underlying data in [S1 Data](#)). (EPS)

## Acknowledgments

We thank Oren Kolodny for commenting on a previous version of this manuscript, and Noa Truskanov for insightful and fruitful discussions.

## Author Contributions

**Conceptualization:** Yosef Prat, Redouan Bshary, Arnon Lotem.

**Formal analysis:** Yosef Prat.

**Investigation:** Yosef Prat, Redouan Bshary, Arnon Lotem.

**Methodology:** Yosef Prat, Redouan Bshary, Arnon Lotem.

**Software:** Yosef Prat.

**Supervision:** Redouan Bshary, Arnon Lotem.

**Visualization:** Yosef Prat.

**Writing – original draft:** Yosef Prat, Redouan Bshary, Arnon Lotem.

**Writing – review & editing:** Yosef Prat, Redouan Bshary, Arnon Lotem.

## References

1. Byrne RW, Bates LA. Sociality, evolution and cognition. *Curr Biol.* 2007; 17:R714–R723. <https://doi.org/10.1016/j.cub.2007.05.069> PMID: 17714665
2. Bluff LA, Weir AAS, Rutz C, Wimpenny JH, Kacelnik A. Tool-related cognition in New Caledonian crows. *Comp Cogn Behav Rev.* 2007; 2:1–25.
3. Emery NJ, Clayton NS. The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science.* 2004; 306:1903–1907. <https://doi.org/10.1126/science.1098410> PMID: 15591194
4. Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. Recursive syntactic pattern learning by songbirds. *Nature.* 2006; 440:1204–7. <https://doi.org/10.1038/nature04675> PMID: 16641998
5. Scarf D, Boy K, Uber Reinert A, Devine J, Güntürkün O, Colombo M. Orthographic processing in pigeons (*Columba livia*). *Proc Natl Acad Sci.* 2016; 113:11272–11276. <https://doi.org/10.1073/pnas.1607870113> PMID: 27638211
6. Murphy RA, Mondragón E, Murphy VA. Rule learning by rats. *Science.* 2008; 319:1849–1851. <https://doi.org/10.1126/science.1151564> PMID: 18369151
7. Harten L, Katz A, Goldshtein A, Handel M, Yovel Y. The ontogeny of a mammalian cognitive map in the real world. *Science.* 2020; 369:194–197. <https://doi.org/10.1126/science.aay3354> PMID: 32647001

8. Tsoar A, Nathan R, Bartan Y, Vyssotski A, Dell’Omo G, Ulanovsky N. Large-scale navigational map in a mammal. *Proc Natl Acad Sci*. 2011; 108:E718–24. <https://doi.org/10.1073/pnas.1107365108> PMID: 21844350
9. Whiten A, Byrne RW. Tactical deception in primates. *Behav Brain Sci*. 2010. 1988; 11:233–244. <https://doi.org/10.1017/S0140525X00049682>
10. Janik VM, Slater PJB. The different roles of social learning in vocal communication. *Anim Behav*. 2000; 60:1–11. <https://doi.org/10.1006/anbe.2000.1410> PMID: 10924198
11. MacLean EL, Matthews LJ, Hare BA, Nunn CL, Anderson RC, Aureli F, et al. How does cognition evolve? Phylogenetic comparative psychology. *Anim Cogn*. 2012; 15:223–38. <https://doi.org/10.1007/s10071-011-0448-8> PMID: 21927850
12. Cauchoix M, Chaine AS. How can we study the evolution of animal minds? *Front Psychol*. 2016; 358. <https://doi.org/10.3389/fpsyg.2016.00358> PMID: 27014163
13. Pinker S, Bloom P. Natural language and natural selection. *Behav Brain Sci*. 1990; 13:707–27. <https://doi.org/10.1017/S0140525X00081061>
14. Berwick RC, Friederici AD, Chomsky N, Bolhuis JJ. Evolution, brain, and the nature of language. *Trends Cogn Sci*. 2013; 17:89–98. <https://doi.org/10.1016/j.tics.2012.12.002> PMID: 23313359
15. Whiten A, editor. *Natural theories of mind: Evolution, development and simulation of everyday mind-reading*. Cambridge, MA, US: Basil Blackwell; 1991.
16. Gallese V, Goldman A. Mirror neurons and the simulation theory of mind-reading. *Trends Cogn Sci*. 1998; 2: 493–501. [https://doi.org/10.1016/s1364-6613\(98\)01262-5](https://doi.org/10.1016/s1364-6613(98)01262-5) PMID: 21227300
17. Rizzolatti G, Craighero L. The mirror-neuron system. *Annu Rev Neurosci*. 2004; 27:169–92. <https://doi.org/10.1146/annurev.neuro.27.070203.144230> PMID: 15217330
18. Solan Z, Horn D, Ruppin E, Edelman S. Unsupervised learning of natural languages. *Proc Natl Acad Sci*. 2005; 102:11629–34. <https://doi.org/10.1073/pnas.0409746102> PMID: 16087885
19. Heyes C. Where do mirror neurons come from? *Neurosci Biobehav Rev*. 2010; 34:575–583. <https://doi.org/10.1016/j.neubiorev.2009.11.007> PMID: 19914284
20. Heyes CM. Theory of mind in nonhuman primates. *Behav Brain Sci*. 1998 Feb 1. 1998; 21:101–114. <https://doi.org/10.1017/s0140525x98000703> PMID: 10097012
21. Galef BG. The question of animal culture. *Hum Nat*. 1992; 3:157–78. <https://doi.org/10.1007/BF02692251> PMID: 24222403
22. Lipkind D, Marcus GF, Bemis DK, Sasahara K, Jacoby N, Takahasi M, et al. Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature*. 2013; 498:104–8. <https://doi.org/10.1038/nature12173> PMID: 23719373
23. Kolodny O, Edelman S, Lotem A. Evolution of protolinguistic abilities as a by-product of learning to forage in structured environments. *Proc R Soc B Biol Sci*. 2015; 282:20150353. <https://doi.org/10.1098/rspb.2015.0353> PMID: 26156764
24. Kolodny O, Lotem A, Edelman S. Learning a generative probabilistic grammar of experience: a process-level model of language acquisition. *Cogn Sci*. 2015; 39:227–267. <https://doi.org/10.1111/cogs.12140> PMID: 24977647
25. Galef BG. Laboratory studies of imitation/field studies of tradition: Towards a synthesis in animal social learning. *Behav Process*. 2015; 112:114–119. <https://doi.org/10.1016/j.beproc.2014.07.008> PMID: 25058622
26. Bellmund JLS, Gärdenfors P, Moser EI, Doeller CF. Navigating cognition: Spatial codes for human thinking. *Science*. 2018; 362:eaat6766. <https://doi.org/10.1126/science.aat6766> PMID: 30409861
27. Skinner BF. The reinforcing effect of a differentiating stimulus. *J Gen Psychol*. 1936; 14:263–78. <https://doi.org/10.1080/00221309.1936.9713154>
28. Williams BA. Conditioned reinforcement: experimental and theoretical issues. *Behav Anal*. 1994; 17:261–85. <https://doi.org/10.1007/BF03392675> PMID: 22478192
29. Kolodny O, Edelman S, Lotem A. The evolution of continuous learning of the structure of the environment. *J R Soc Interface*. 2014; 11:20131091–1. <https://doi.org/10.1098/rsif.2013.1091> PMID: 24402920
30. Enquist M, Lind J, Ghirlanda S. The power of associative learning and the ontogeny of optimal behaviour. *R Soc Open Sci*. 2019; 3:160734. <https://doi.org/10.1098/rsos.160734> PMID: 28018662
31. Leadbeater E. What evolves in the evolution of social learning? *J Zool*. 2015; 295:4–11. <https://doi.org/10.1111/jzo.12197>
32. Epstein RA, Patai EZ, Julian JB, Spiers HJ. The cognitive map in humans: spatial navigation and beyond. *Nat Neurosci*. 2017; 20:1504–13. <https://doi.org/10.1038/nn.4656> PMID: 29073650

33. Troyer TW, Doupe AJ. An associational model of birdsong sensorimotor learning I. Efference copy and the learning of song syllables. *J Neurophysiol.* 2000; 84:1204–23. <https://doi.org/10.1152/jn.2000.84.3.1204> PMID: 10979996
34. Abe K, Watanabe D. Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat Neurosci.* 2011; 14:1067–74. <https://doi.org/10.1038/nn.2869> PMID: 21706017
35. Clayton NS, Bussey TJ, Dickinson A. Can animals recall the past and plan for the future? *Nat Rev Neurosci.* 2003; 4:685–91. <https://doi.org/10.1038/nrn1180> PMID: 12894243
36. Klump BC, Martin JM, Wild S, Hörsch JK, Major RE, Aplin LM. Innovation and geographic spread of a complex foraging culture in an urban parrot. *Science.* 2021; 373:456–60. <https://doi.org/10.1126/science.abe7808> PMID: 34437121
37. Auersperg AMI, Kacelnik A, von Bayern AMP. Explorative learning and functional inferences on a five-step means-means-end problem in Goffin's Cockatoos (*Cacatua goffini*). *PLoS ONE.* 2013; 8:e68979. <https://doi.org/10.1371/journal.pone.0068979> PMID: 23844247
38. Pearce JM. Evaluation and development of a connectionist theory of configural learning. *Anim Learn Behav.* 2002; 30:73–95. <https://doi.org/10.3758/bf03192911> PMID: 12141138
39. Duncan K, Doll BB, Daw ND, Shohamy D. More than the sum of its parts: a role for the hippocampus in configural reinforcement learning. *Neuron.* 2018; 98:645–657.e6. <https://doi.org/10.1016/j.neuron.2018.03.042> PMID: 29681530
40. Gobet F, Lane P, Croker S, Cheng P, Jones G, Oliver I, et al. Chunking mechanisms in human learning. *Trends Cogn Sci.* 2001; 5:236–43. [https://doi.org/10.1016/s1364-6613\(00\)01662-4](https://doi.org/10.1016/s1364-6613(00)01662-4) PMID: 11390294
41. Jones G. Why chunking should be considered as an explanation for developmental change before short-term memory capacity and processing speed. *Front Psychol.* 2012; 3:1–8. <https://doi.org/10.3389/fpsyg.2012.00001> PMID: 22279440
42. Brent MR. Speech segmentation and word discovery: A computational perspective. *Trends Cogn Sci.* 1999:294–301. [https://doi.org/10.1016/s1364-6613\(99\)01350-9](https://doi.org/10.1016/s1364-6613(99)01350-9) PMID: 10431183
43. Goldstein MH, Waterfall HR, Lotem A, Halpern JY, Schwade J, et al. General cognitive principles for learning structure in time and space. *Trends Cogn Sci.* 2010; 14:249–58. <https://doi.org/10.1016/j.tics.2010.02.004> PMID: 20395164
44. Whitlow JW, Wagner AR. Negative patterning in classical conditioning: Summation of response tendencies to isolable and configural components. *Psychon Sci.* 1972; 27:299–301. <https://doi.org/10.3758/BF03328970>
45. Rescorla RA, Wagner AR. A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF, editors. *Classical conditioning II: current research and theory.* Appleton-Century-Crofts; 1972. p. 64–99.
46. Mackintosh NJ. A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychol Rev.* 1975; 82:276–98. <https://doi.org/10.1037/h0076778>
47. Sutherland RJ, Rudy JW. Configural association theory: The role of the hippocampal formation in learning, memory, and amnesia. *Psychobiology.* 1989; 17:129–44. <https://doi.org/10.3758/BF03337828>
48. Papadimitriou A, Wynne CDL. Preserved negative patterning and impaired spatial learning in pigeons (*Columba livia*) with lesions of the hippocampus. *Behav Neurosci.* 1999; 113:683–90. <https://doi.org/10.1037/0735-7044.113.4.683> PMID: 10495077
49. Fournier DI, Todd TP, Bucci DJ. Permanent damage or temporary silencing of retrosplenial cortex impairs the expression of a negative patterning discrimination. *Neurobiol Learn Mem* 2019; 163:107033. <https://doi.org/10.1016/j.nlm.2019.107033> PMID: 31173918
50. Conway CM. How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neurosci Biobehav Rev.* 2020; 112:279–99. <https://doi.org/10.1016/j.neubiorev.2020.01.032> PMID: 32018038
51. Lotem A, Halpern JY. Coevolution of learning and data-acquisition mechanisms: a model for cognitive evolution. *Philos Trans R Soc Lond B Biol Sci.* 2012; 367:2686–94. <https://doi.org/10.1098/rstb.2012.0213> PMID: 22927567
52. Lotem A, Halpern JY, Edelman S, Kolodny O. The evolution of cognitive mechanisms in response to cultural innovations. *Proc Natl Acad Sci.* 2017; 114:7915–7922. <https://doi.org/10.1073/pnas.1620742114> PMID: 28739938
53. Côté IM. Evolution and ecology of cleaning symbioses in the sea. In: Gibson RN, Barnes M, editors. *Oceanography and Marine Biology: An Annual Review: Volume 38: An Annual Review.* CRC Press; 2000. p. 311–355.
54. Bshary R. The cleaner fish market. In: Noë R, Hooff JARAM, Hammerstein P, editors. *Economics in Nature.* Cambridge University Press; 2001. p. 146–172. <https://doi.org/10.1098/rspb.2001.1495> PMID: 11454294

55. Bshary R, Noë R. Biological markets: the ubiquitous influence of partner choice on the dynamics of cleaner fish-client reef fish interactions. In: Hammerstein P, editor. Genetic and cultural evolution of cooperation. MIT press; 2003. p. 167–184. <https://doi.org/10.1098/rsbl.2003.0077> PMID: 14667394
56. Triki Z, Wismer S, Rey O, Ann Binning S, Levorato E, Bshary R. Biological market effects predict cleaner fish strategic sophistication. *Behav Ecol*. 2019; 30:1548–57. <https://doi.org/10.1093/beheco/arz111>
57. Bshary R. Machiavellian intelligence in fishes. *Fish Cognition and Behavior*; 2011. p. 277–297. <https://doi.org/10.1002/9781444342536.ch13>
58. Wismer S, Pinto AI, Vail AL, Grutter AS, Bshary R. Variation in cleaner wrasse cooperation and cognition: influence of the developmental environment? *Ethology*. 2014; 120:519–531. <https://doi.org/10.1111/eth.12223>
59. Bshary R, Grutter AS. Experimental evidence that partner choice is a driving force in the payoff distribution among cooperators or mutualists: the cleaner fish case. *Ecol Lett*. 2002; 5:130–136. <https://doi.org/10.1046/j.1461-0248.2002.00295.x>
60. Truskanov N, Emery Y, Bshary R. Juvenile cleaner fish can socially learn the consequences of cheating. *Nat Commun*. 2020; 11:1159. <https://doi.org/10.1038/s41467-020-14712-3> PMID: 32127522
61. Triki Z, Wismer S, Levorato E, Bshary R. A decrease in the abundance and strategic sophistication of cleaner fish after environmental perturbations. *Glob Chang Biol*. 2018; 24:481–9. <https://doi.org/10.1111/gcb.13943> PMID: 29134754
62. Salwiczek LH, Prétôt L, Demarta L, Proctor D, Essler J, Pinto AI, et al. Adult cleaner wrasse outperform capuchin monkeys, chimpanzees and orang-utans in a complex foraging task derived from cleaner–client reef fish cooperation. *PLoS ONE*. 2012; 7:e49068. <https://doi.org/10.1371/journal.pone.0049068> PMID: 23185293
63. Staddon JER, Cerutti DT. Operant conditioning. *Annu Rev Psychol*. 2003; 54:115–44. <https://doi.org/10.1146/annurev.psych.54.101601.145124> PMID: 12415075
64. Prétôt L, Bshary R, Brosnan SF. Factors influencing the different performance of fish and primates on a dichotomous choice task. *Anim Behav*. 2016; 119:189–199. <https://doi.org/10.1016/j.anbehav.2016.06.023>
65. Zentall TR, Case JP, Luong J. Pigeon's (*Columba livia*) paradoxical preference for the suboptimal alternative in a complex foraging task. *J Comp Psychol*. 2016; 130:138–44. <https://doi.org/10.1037/com0000026> PMID: 27064201
66. Zentall TR, Case JP, Berry JR. Rats' acquisition of the ephemeral reward task. *Anim Cogn*. 2017; 20:419–25. <https://doi.org/10.1007/s10071-016-1065-3> PMID: 27988824
67. Zentall TR, Case JP. The ephemeral-reward task: optimal performance depends on reducing impulsive choice. *Curr Dir Psychol Sci*. 2018; 27:103–9. <https://doi.org/10.1177/0963721417735522>
68. Prétôt L, Mickelberg J, Carrigan J, Stoinski T, Bshary R, Brosnan SF. Comparative performance of orangutans (*Pongo spp.*), gorillas (*Gorilla gorilla gorilla*), and drills (*Mandrillus leucophaeus*), in an ephemeral foraging task. *Am J Primatol*. 2020:e23212. <https://doi.org/10.1002/ajp.23212> PMID: 33135209
69. Pepperberg IM, Hartsfield LA. Can Grey parrots (*Psittacus erithacus*) succeed on a “complex” foraging task failed by nonhuman primates (*Pan troglodytes*, *Pongo abelii*, *Sapajus apella*) but solved by wrasse fish (*Labroides dimidiatus*)? *J Comp Psychol*. 2014; 128:298–306. <https://doi.org/10.1037/a0036205> PMID: 24798239
70. Zentall TR. The paradoxical performance by different species on the ephemeral reward task. *Learn Behav*. 2020. <https://doi.org/10.3758/s13420-020-00429-2> PMID: 32583140
71. Zentall TR, Case JP, Berry JR. Early commitment facilitates optimal choice by pigeons. *Psychon Bull Rev*. 2017; 24:957–63. <https://doi.org/10.3758/s13423-016-1173-8> PMID: 27743217
72. Quiñones AE, Leimar O, Lotem A, Bshary R. Reinforcement learning theory reveals the cognitive requirements for solving the cleaner fish market task. *Am Nat*. 2020; 195:664–77. <https://doi.org/10.1086/707519> PMID: 32216674
73. Bush RR, Mosteller F. A mathematical model for simple learning. *Psychol Rev*. 1951:313–23. <https://doi.org/10.1037/h0054388> PMID: 14883244
74. McNamara JM, Houston AI. Memory and the efficient use of information. *J Theor Biol*. 1987; 125:385–395. [https://doi.org/10.1016/s0022-5193\(87\)80209-6](https://doi.org/10.1016/s0022-5193(87)80209-6) PMID: 3657218
75. Kelleher RT, Gollub LR. A review of positive conditioned reinforcement. *J Exp Anal Behav*. 1962; 5:543–97. <https://doi.org/10.1901/jeab.1962.5-s543> PMID: 14031747
76. Lotem A, Halpern JY. A data-acquisition model for learning and cognitive development and its implications for autism. *Computing and Information Science Technical Reports*. Cornell University; 2008.

77. Kolodny O, Edelman S, Lotem A. Evolved to adapt: A computational approach to animal innovation and creativity. *Curr Zool.* 2015; 61:350–68. <https://doi.org/10.1093/czoolo/61.2.350>
78. Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. *Science.* 1996; 274:1926–1928. <https://doi.org/10.1126/science.274.5294.1926> PMID: 8943209
79. Byrne RW. Imitation without intentionality. Using string parsing to copy the organization of behaviour. *Anim Cogn.* 1999; 2:63–72. <https://doi.org/10.1007/s100710050025>
80. Sutton RS, Barto AG. Reinforcement learning: An introduction. 2nd ed. MIT press; 2018.
81. Niv Y, Joel D, Meilijson I, Ruppin E. Evolution of reinforcement learning in uncertain environments: a simple explanation for complex foraging behaviors. *Adapt Behav.* 2002; 10:5–24. <https://doi.org/10.1177/1059-712302-010001-01>
82. Lotem A, Nakamura H, Zahavi A. Constraints on egg discrimination and cuckoo-host co-evolution. *Anim Behav.* 1995; 49:1185–1209. <https://doi.org/10.1006/anbe.1995.0152>
83. Rodríguez-Gironés MA, Lotem A. How to detect a cuckoo egg: a signal-detection theory model for recognition and learning. *Am Nat.* 1999; 153:633–48. <https://doi.org/10.1086/303198> PMID: 29585642
84. Miyashita Y. Cognitive memory: cellular and network machineries and their top-down control. *Science.* 2004; 306:435–40. <https://doi.org/10.1126/science.1101864> PMID: 15486288
85. Perruchet P, Vinter A. PARSER: A model for word segmentation. *J Mem Lang.* 1998; 39:246–263. <https://doi.org/10.1006/jmla.1998.2576>
86. Truskanov N, Emery Y, Porta S, Bshary R. Configural learning by cleaner fish in a complex biological market task. *Anim Behav.* 2021; 181:51–60. <https://doi.org/10.1016/j.anbehav.2021.08.023>
87. Triki Z, Levorato E, McNeely W, Marshall J, Bshary R. Population densities predict forebrain size variation in the cleaner fish *Labroides dimidiatus*. *Proc R Soc B Biol Sci.* 2019; 286:20192108. <https://doi.org/10.1098/rspb.2019.2108> PMID: 31744435
88. Triki Z, Emery Y, Teles MC, Oliveira RF, Bshary R. Brain morphology predicts social intelligence in wild cleaner fish. *Nat Commun.* 2020; 11:6423. <https://doi.org/10.1038/s41467-020-20130-2> PMID: 33349638