Andrés E. Quiñones[[1]](#footnote-1)

C. Daniel Cadena\*

Redouan Bshary[[2]](#footnote-2)

March 01, 2023

Adaptive behavioural responses often depend on qualities of the interacting partner of an individual. For example, when competing for resources, an individual might be better off escalating fights with individuals of lower quality, while restraining from fighting individuals of higher quality. Communication systems involving signals of quality allow individuals to reduce uncertainty regarding the fighting ability of their partners and make more adaptive behavioral decisions. However, dishonest individuals can destabilize such communications systems. One open question is whether cognitive mechanisms, such as learning, can maintain the honesty of signals, thus, favoring their evolutionary stability. Here we present evolutionary simulations, where individuals can produce a signal proportional to their quality; and learn along their lifetime the best response to the signal emitted by their peers. In these simulations, learning on the receiver side can mediate the evolution of signals of quality on the sender side. When the cost of the signal is proportional to the quality of the sender populations are only composed of honest signalers. When the cost is not proportional to the quality of the signaler, the population is composed of both honest and dishonest signalers. We argue that learning can be a general cognitive mechanism playing a role in a wide range of communication systems.

## Introduction

The outcome of interactions with con-specifics is an important determinant of fitness in social animals. Irrespective of whether such interactions are cooperative or competitive, the actions of interacting individuals partly determine their reproductive success. However, the best action for an individual might vary depending on its own condition, and that of individuals with whom it interacts. Thus, acting based on information about interacting partners is typically adaptive (Quiñones et al. 2016). Acquiring this information, however, is far from trivial. In some cases, interacting partners (signallers) might be ‘willing’ to provide accurate information, but in others it might be in their own interest to conceal information (Johnstone 1997), or to provide erroneous information (Johnstone and Norris 1993). Typically, in a given context, some proportion of individuals will benefit from broadcasting accurate information, whereas others will benefit from hiding it. Take, for instance, an interaction between two individuals where one can help the other. Given some costs and benefits, a potential donor will be interested in helping related individuals. Therefore, for a relative of the donor, broadcasting kinship is advantageous, whereas for an unrelated individual, concealing the lack of kinship is better. Similar scenarios may apply to many aspects of social life such as finding mates, feeding offspring, dominance relationships and aggressive interactions (Bradbury and Vehrencamp 2011; Møller 1988; Tibbetts and Dale 2004). In dominance and aggressive interactions, an important piece of information to guide individual actions is interacting partners fighting ability, which is sometimes referred to as Resource Holding Potential (RHP) (Parker 1974), or simply as quality. Responsiveness to the quality of the partner is central to communication systems such as badges of status, where an arbitrary signal conveys quality (Johnstone and Norris 1993; Rohwer 1975). Here, arbitrary refers to a signal which is not ontogenetically correlated with quality. Badges of status can be evolutionarily stable whenever the signal imposes a fitness cost which decreases with the quality of an individual (Botero et al. 2010; Johnstone and Norris 1993), such that it’s no longer in the interest of low quality individuals to fake quality by producing a large badge. The cost of honest signals is a more general principle of communications systems, usually referred to as the handicap principle (Grafen 1990; Zahavi 1975).

A somewhat ignored component of communication systems, in the context of aggressive interactions, is the cognitive aspects of the receiver module. Theoretical models often assume that communication systems relying on badges of status to have reaction norms as their mechanistic underpinning (Botero et al. 2010), with individuals using their own quality and the opponent’s badge to determine whether to aggressively engage in contest. Reaction norms allow individuals to respond to the available information without a big cognitive burden. That is at least in contrast to other information processing mechanisms like individual recognition. With individual recognition individuals associate cue of their peers to their quality. These association must be learned throughout an individual’s life, thus it has the usual cognitive requirements of an associative learning process. An alternative view of badges of status is that individuals learn to react to them based on their experiences (Guilford and Dawkins 1991), which would imply that receivers must learn to associate signals with the fighting ability of bearers just as in systems based on individual recognition. In cases where badges of status vary quantitatively (e.g. in size or intensity), fighting ability may increase monotonically with attributes of the badge and would be reinforced by every interaction, so in principle learning would be faster than in systems involving individual recognition in which the association between signals and their meaning would vary depending on the interacting partners. In any case, learning could be a central cognitive mechanism in both types of communication systems, but the role of learning in these contexts has not been thoroughly explored in the empirical nor theoretical literature.

Associative learning is a key cognitive mechanism that allows individuals to associate rewards with environmental stimuli and thus behave adaptively (Staddon 2016). Associative learning exists in all major vertebrate taxa, and in many invertebrates as well (Heyes 2012; Macphail 1982; Staddon 2016; Behrens et al. 2008). Theory has shown that natural selection favours these associations in complex environments where conditions are hard to predict (Dridi and Lehmann 2016). Besides its wide taxonomic and ecological relevance, associative learning is a flexible cognitive mechanism whose underpinnings show interspecific variation (Enquist, Lind, and Ghirlanda 2016; Quiñones et al. 2019). Associative learning is not a single mechanisms, but rather a set of cognitive structures and processes that can vary in their scope and complexity. Presumably, these structures and processes have been modified by natural selection and could provide new explanations for behavioural variation. Despite all this, associative learning is not often included in the narrative of evolutionary explanations of behavioural patterns (Fawcett, Hamblin, and Giraldeau 2013; Kamil 1983; McAuliffe and Thornton 2015). Computational models of evolution can play an important role to overcome the lack of integration between learning and evolution. Reinforcement learning theory encompasses a series of computational methods inspired on the psychological and neurological mechanisms of associative learning (Sutton and Barto 2018). These set of algorithms allow the implementation of biologically realistic problems, capturing the essence of learning processes (Frankenhuis, Panchanathan, and Barto 2018; Quiñones et al. 2019). Furthermore, these algorithms can be embedded in evolutionary simulations to generate theoretical predictions of the effect that learning can have in behavioural evolution (Leimar and McNamara 2019; Leimar and Bshary 2022b).

In here we present an evolutionary model where individuals use associative learning to develop a tendency to behave aggressively (or peacefully), in the context of competition over resources, depending on the quantitative morphological trait they perceive in their opponent (badge). Over evolutionary time individuals evolve the size of their badge, and whether it depends on their quality. Under this simple set up, individuals can use the badge as a signal of quality. We use the model to assess under what conditions of interaction structure we expect different communication signals to evolve.

## The model

We model the evolution of signals of quality in the context of agonistic interactions. For simplicity we consider a population of haploid individuals with non-overlapping generations. Individuals are born every generation with a level of quality (, where is subscript of the population vector of size ) given by a number between zero and one, which is drawn from a truncated normal distribution . An individual with quality has the lowest RHP, while an individual with quality has the highest. As part of development processes individuals produce a morphological signal (badge), the size of such depends on their level of quality according to a reaction norm given by ; where and are individual specific traits that determine the shape of the reaction norm (Fig. A). Variation in and means individuals can have either uninformative (flat) or informative (logistic) norms. The size of the signal is constrained between zero and one. We assume different values of these traits are given by different alleles, and are inherited from mother to offspring. Unless, with a small probability (), mutation changes the allelic value of the offspring by an amount taken from a normal distribution . After birth individuals go through a round of viability selection. Individual specific survival probability is given by . Where and are parameter values determining the shape of the survival function. Importantly, if survival probability is independent of quality, while if lower quality individuals pay a higher price for similar-sized badge, fulfilling the core assumption of the handicap principle (Botero et al. 2010; Grafen 1990; Johnstone and Norris 1993).

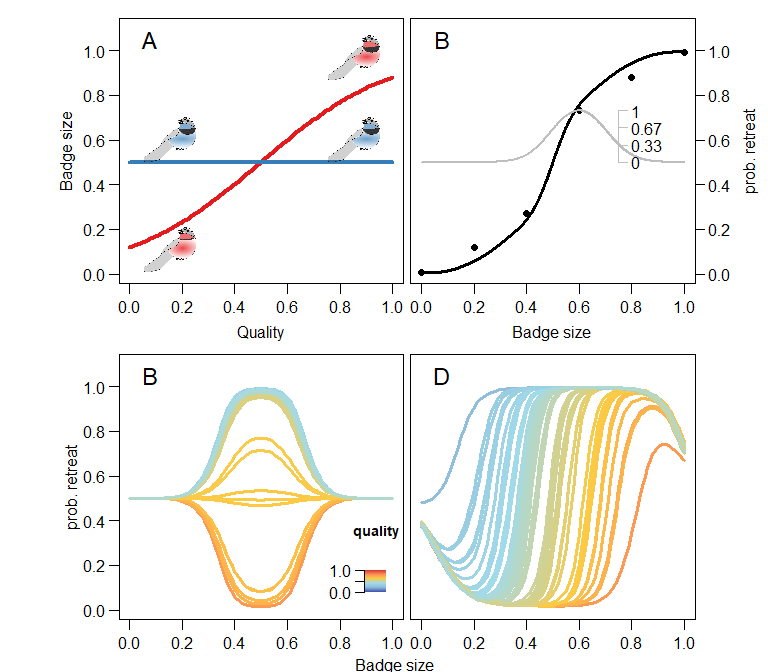
Individuals who survive engage in a series of pairwise interactions where they compete for resources. In an interaction each individual must decide whether to escalate a fight or not. Following the classic *hawk-dove* game (Maynard-Smith 1982), if the focal individual fights and its partner does not, the focal gets as pay-off the resource of value while its partner gets nothing; if both individuals restrain from fighting they split up the resource in half; if both decide to fight they split up the cost of fighting and the winner takes over the resource. We further assume that the probability of wining a fight for the focal individual depends on the difference in quality between her and her interacting partner; specifically it is given by , where is a parameter defining how strong the quality difference determines the winning probability; and and denote the the position in the population vector of the focal and its interacting partner respectively.

The decision of whether to escalate a fight against an interacting partner can be dependent on the size of the partner’s badge, the actual dependency is determined by the focal´s experiences through a learning process. We implement learning using the actor-critic approach from reinforcement learning (RL) theory (Sutton and Barto 2018; Quiñones et al. 2019; Leimar and McNamara 2019). Individuals estimate the reward (pay-off) expected from interacting with partners of different badge sizes (the critic in RL terminology). After every interaction they update the estimate of reward proportionally to the difference between their current estimate and the observed reward (prediction error ) and to the speed of learning (). Furthermore, they express different probabilities of retreating (attacking) depending on the badge size of their opponent (the actor in RL). They update the probability of retreating (attacking) up or down depending on whether retreating (attacking) leads to an increase in the reward estimation. So, if a focal individual decides to escalate a fight against an individual with small badge, and that leads to an increase in the reward estimation, the focal individual will increase the probability of escalating fights with individuals of small badges in the future. Given that badge size is a real number between 0 and 1, there are infinitely many badge sizes. Thus, the reward estimation, as well as the probability of retreating (escalating), must be generalized across different values. To implement generalization we use the linear function approximation method based on radial basis functions (Sutton and Barto 2018). Specifically, we pick feature centres, which are evenly spaced values along the badge size interval ([0,1]). These feature centres are the same for all individuals and stored in vector . Each one of these features centres is associated with a weight for reward estimation and tendency to play retreat in a given interaction. The reward estimation and the tendency to retreat are calculated (in every interaction) as the sum of the weights associated with each feature centre, weighted by the response triggered by the feature (Fig. B, dots represent the feature weights) The response of each feature centre diminishes as a Gaussian function with the distance between the feature centre and the badge size of the partner (Fig. B, grey line). Formally the reward estimation when the focal individual () faces individual is given by,

where is the weight of feature centre on the reward estimation; and is the width of the generalization function. Similarly, the tendency to retreat is given by the sum of feature weights associated with the actor, and the probability is obtained by applying a logistic transformation (Fig. B, black line). So formally, the log-odds to retreat when facing individual is

where is the weight of feature centre on the tendency to retreat.

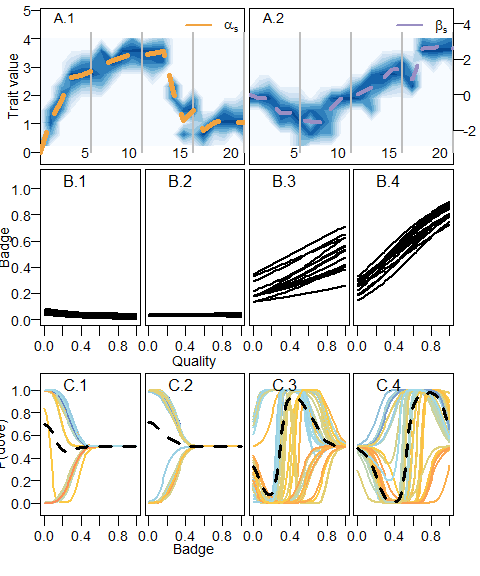
After all their interactions, individuals in the population reproduce proportionally to their total pay-off , which is a sum of the baseline pay-off () and all the pay-offs obtained throughout their life. Thus, the combination of natural selection and genetic drift changes the distribution of values in and that segregates in the population, effectively changing the badges expressed and the communication system.



Model of communication in the context of aggressive interactions. In A the sender code: a reaction norm determines how the badge size is determined by the quality of the individual. Red shows an informative reaction norm, while the blue shows a uninformative reaction norm. B the receiver code: individuals have a behavioral reaction norm that determines their probability of retreating (black line). The reaction norm arises from generalizing the information from the feature weights (black dots). Generalization is represented by the grey line and its axis, which shows the response triggered by the fourth feature diminishes as the value evaluated is further from the feature center. The learning process moves the feature weights (black dots) up or down depending on whether that leads to an increase in the estimated reward. In C and D, effect of learning on the receiver strategy. Receivers in C face signallers with uninformative reaction norms (blue line in A), while in D they face signallers with informative reaction norms (red line in A). Colour scale in C (applies also for D) indicates the quality of the individual

## Results

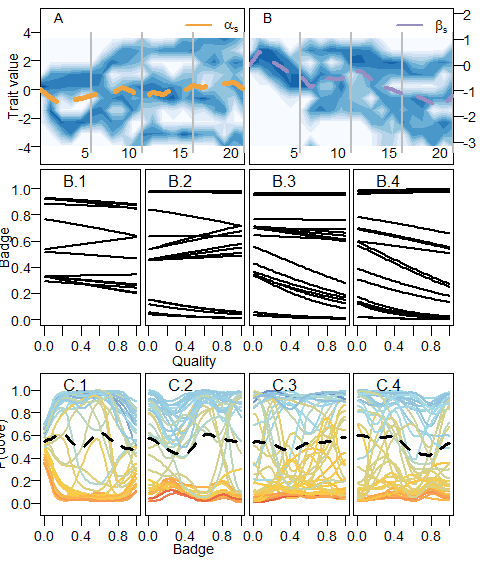
We first present the outcome of simulations where we prevent the evolution of the sender code (by setting the mutation rate to ), and assume that all individuals in the population display either an uninformative or informative badge size with respect to their quality. These simulations show what type of receiver strategy develops through a learning process. Figure (C and D) shows the receiver strategy developed through learning; panel C is for receivers that faced uninformative signals, and D informative ones. When learners face uninformative signals, they change their probability of retreating depending on their own quality. Individuals with higher quality (red tones) after the learning process have a high probability of attacking; while individuals of lower quality (blue tones) mostly retreat from confrontations. Thus, learning splits the population of receivers into the two classic pure strategies of hawks and doves. Given that we have a assumed a monomorphic population with unresponsive reaction norms on the signalling side, the changes triggered by learning only affect a small range of badge sizes. In contrast, when receivers face informative reaction norms on the side of the signaller (panel B in fig. ), receivers use the badge size of their interacting partners to determine whether to retreat or attack. The relationship is given by a threshold-like reaction norm, where the change from retreating to attacking depends on the quality of the receiver. As expected, the higher the quality of the receiver the larger the badge size that triggers a retreat.

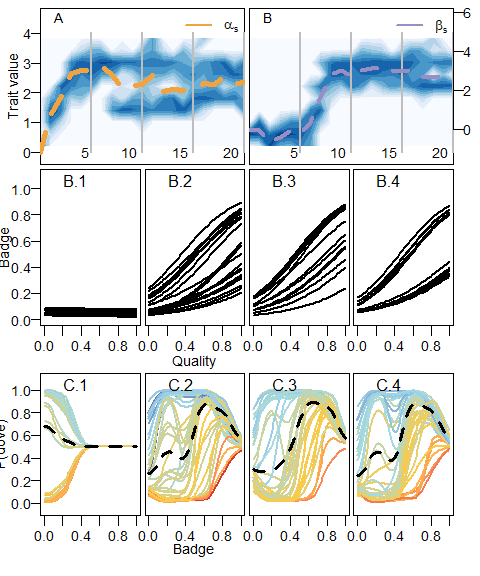


Evolution of the badge size as a handicap mediated by learning. Top panels (A.1 and A.2) show the evolutionary dynamics of the sender code. On the left (A.1), changes in the distribution of values for the intercept of the reaction norm (); on the right (A.2) changes in the distribution of the slope (). Darker areas of the background correspond to values with high frequency, while dashed lines in both panels show the mean of the distribution. X axis in the evolutionary dynamics are given in thousands of generations. Grey lines show the generation time corresponding to the panels below portraing the sender (B) and receiver code(C). In the bottom panels (C1-4), the learned reaction norms correspond to individuals in the population after the interaction round. Color scale indicates quality just as in Fig. .

When we allow the evolution of the sender code ( and change subject to natural selection and genetic drift), and the badge size work as a handicap (i.e. the cost of the badge is inversely proportional to the quality of the individual), the sender code can evolve to produce an honest signal (Fig. ). The evolution of the sender code does not happen immediately after the start of the evolutionary process. Instead, the evolutionary dynamics of the reaction norm parameters (Fig. , A) show a set of steps. First, the intercept of the reaction norm () evolves to higher values, reducing the average badge size in the populations. That makes sense because reaction norms segregating in the population are flat at this point, and consequently the receiver code does not respond to the badge size, bigger badges are costlier and do not trigger lower attack probabilities. During those first generations, the slope of the sender reaction norm () remains around the initial value of zero. At around generation 4000 in our simulations, the slope evolves toward positive values. Positive values in the slope imply that larger badges correlate with higher quality (Fig. B). Receivers learn to react to such correlation, increasing the probability of retreat when facing individual of larger badges (Fig. C). Hence, natural selection favours larger values of , eventually leading to an evolutionary equilibrium where badge size is an honest signal of quality mediated by the learned responses of receivers (Fig. B.4 and C.4). The evolutionary trajectory portrayed in figure A is not the only possible trajectory. If the slope of the sender reaction norms () evolves toward negative values before badges become handicaps, receivers never learn to react to the size of badges. Thus, badges are only costly and do not provide information about the quality of an individual (Fig. ) . Eventually, the badge disappears from the population. In contrast, when the badge does not work as a handicap but instead is cost-free, the evolutionary process never leads to the establishment of an honest signal. Instead, subject to genetic drift the population either evolves towards the disappearance of the badge, or to the maximum badge size. In either of this cases reaction norms are flat, so the badge does not provide any quality information.

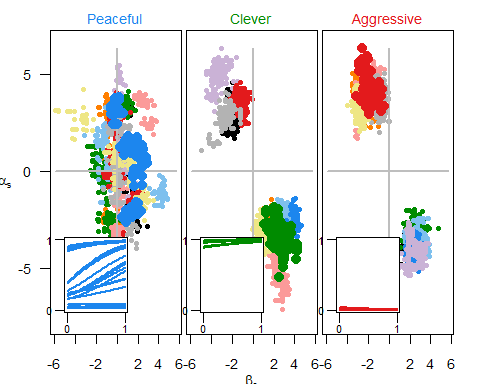
The amount of information that agents are able to collect through learning throughout their life time can strongly change the outcome of the evolutionary dynamics. In the simulations presented in figure and , individuals learned with high learning rate () and interacted repeatedly along their lifetime (on average 2000 interactions). When we reduce the number of interactions that individuals have over their lifetime to 300 interactions on average, we see an drastic increase in the morphological and genetic variation present in the population. Fig. shows changes in the distribution of the sender reaction norm parameters along evolutionary time, assuming the badge is cost-free. In panels A1 and 2, darker areas show trait values with high frequency in the population. Populations start monomorphic with a value of zero on both traits, and mutations quickly build up a normal distribution around the starting value. Within the first 2000 generations the unimodal distribution in the intercept () splits into a bimodal one. Later in evolutionary time one of the peaks splits further, so at the end of the evolutionary simulation the distribution of the intercept () in the population shows three peaks. In the case of , the peaks in the distribution are not so clear-cut, but it’s clear that there is an increase in the variance of the distribution. These changes in the distribution of the sender reaction norm parameters imply that individuals can generally be classified into three distinct types (). Two types express a flat reaction norm with extreme values for the badge size, meaning that their badge size is not informative of their quality, while the third type shows intermediate badge sizes which are determined by the quality of the individual (Fig. B.3-4). In this particular simulation, the values of the slope in the informative group are negative, which implies that badge size correlates negatively with quality. Furthermore, the receiver reaction norms developed through learning, particularly those of individuals of intermediate quality, respond to the signal of their sender type by increasing the probability of retreating from a fight with individuals with smaller badges (Fig. ).

 The effect of a limited number of interactions on the amount of variation in the evolving parameters plays a role when the signal follows the handicap principle as well. In fig , we show simulations where individuals have on average 300 interactions in their life and the cost of the signal is proportional to the quality, following the handicap principle. The evolutionary process leads to a combination of reaction norm parameters where there is a positive correlation between the size of the badge and the quality of the individual. This relationship, however, is muddled by the fact that there are two clusters of values for the intercept () and slope () of the reaction norm in the population. Thus, there are two types of reaction norms. One with a steeper slope, meaning that it expresses a higher badge size for a given quality. This effect of an increased variance in the trait distribution is not only triggered by lower number of interactions. Larger variances are also found when we assume a lower speed of learning (data not shown). This suggest that limits to the amount of information that individuals acquire through learning allows the coexistence of different communication strategies (Botero et al. 2010).



The evolution of diverse costly signals. Portrait of the evolutionary dynamics of the sender code with snapsots of both sender and receiver codes just as in fig. . The middle panels show changes in the distribution of values for and along evolutionary time. Panels above and below correspond to snapshots of the sender and receiver codes, respectively, generation time of the snapshots are indicated by the grey lines in the middle panels.

The behaviour expressed by naive individuals (those who have not learned) imposes negative-frequency dependent selection, allowing for the build up of genetic variation in reaction norms. In the simulations presented so far, we assumed that individuals start with a flat behavioural reaction norm such that they escalate aggressively a fight with a probability regardless of the badge size of the interacting partner. To asses if this initial conditions of the communication system had any role in the build up of genetic variation, we ran a series of simulations varying the initial conditions of the actor module in the learning model. Specifically, we let naive individuals have a flat reaction norm with a 1) low (peaceful), 2) high (aggressive) probability of escalating a fight and 3) a probability corresponding to the ESS of the original hawk-dove game (clever). Fig. shows the distribution of values of the intercept () and slope () evolved in different replicates of the simulations. The left-hand side panel, that corresponds to peaceful naive individuals, is the only one where the distribution of values is split in different clusters. That is, the majority of the variation occurs within clusters. In contrast, when naive individuals behave either aggressively or cleverly, the badges evolve toward minimum and maximum values, and so the variation is driven mainly by among-replicate differences. We can make sense of these results by realizing that individuals change their naive behaviour quickly in ranges of the badge size that are common in the population. Thus, one way to repeatedly interact with the naive behaviour is to have a rare badge size. When the naive behaviour is peaceful, individual with a rare badge size have a fitness advantage. This triggers negative-frequency dependent selection and the evolution of different types of badge sizes. In situations where individuals learn fast and interact repeatedly the strength of frequency dependent selection diminishes. The diversification process presented here is fully dependent on the learning process and how this process is implemented.



The peaceful, the clever and the aggressive. Distribution of values of the intercept and slope for individuals at the end of the evolutionary simulations. The panels show the three different initial conditions for the behaviour of individuals: peaceful (blue), clever (green) and aggressive (red), see maintext for details. Colours indicate the replicate simulation. In the inset, the resulting reaction norms corresponding to the intecept and slope values for one of the replicates, which replicate is shown is indicated by the line colour.

## Discussion

We have presented here an evolutionary model where the response of a signal in a communication system is mediated by learning processes. Individuals throughout their life learn the best way to respond to a quantitative trait of their interaction partners. The learning process can mediate the evolution of an honest signal under the handicap principle (Grafen 1990; Zahavi 1975). This conclusion is not different from the classical genetic models, where responses to the signal are inherent to the individual. However, unlike classical genetic models, our model shows that learning can also mediate a signal polymorphism in the absence of the handicap principle. Under this polymorphic equilibrium different types of individuals, with honest and dishonest signals, compose the population. The amount of information gathered through learning, as a well as the initial conditions of the learning process, are crucial for the emergence and maintenance of this variation. The emergence of the polymorphism requires a limit in the amount of information collected by individuals, as well as peaceful behaviour in naive individuals.

Associative learning is a powerful mechanism to learn about the world, about social partners and about one’s abilities relative to the social context. By associating cues and signals with fitness relevant outcomes individuals collect information that allows them to improve their reproductive potential. This is very obvious in classical examples such as when animals learn to avoid food that makes them sick, or use environmental cues to find food. More recent ideas have highlighted the potential role that associations can have in a social contexts such as: hierarchy formation (Leimar and Bshary 2022a, 2022b), and cooperation (Leimar and McNamara 2019; Dridi and Akçay 2018). In this theoretical examples individuals use various sources of reward to make adaptive choices in a given social context. Here we extend this logic to the evolution of a communication system mediated by a quality signal. Previous evolutionary models of badges of status assumed that individuals responded using a behavioural reaction norm, where the opponent’s badge and individual’s own quality determined the behaviour (Botero et al. 2010). However, there isn’t a clear mechanism to justify the assumption that individuals inherently know their own quality, particularly relative to their peers. In our model, individuals not only learn about the quality signal, but also learn about their own quality relative to others. This can be seen throughout our simulations in the responses developed by individuals of different quality. Because individuals learn about their own quality, even in the absence of an honest signal, they are able to make adaptive choices. Thus the learning process we have modelled and the response it mediates has fitness relevant consequences even in the absence of an honest signal.

Learning processes collecting information on a population-wide level promote frequency dependent selection, in the communication system we model frequency dependency translates to the coexistence of different communication strategies. The frequency-dependence triggered by learning process was highlighted by the classic study of Bond and Kamil (2002) where live predators drive the evolution of *in silico* polymorphic prey. The key to the evolution of this polymorphisms is that predators learn to better discriminate preys that are frequently encountered. Thus, preys found in low frequency have a selective advantage. An analogous process occurs in our simulations, the learning algorithm of the receiver individual learns the best response for herself towards individuals that have a badge size common in the population. When an individual has a rare badge size it can be favoured or unfavoured by the naive behaviour. When the naive behaviour is peaceful, polymorphism is promoted, when the naive behaviour is aggressive polymorphism is prevented. These two examples show how behaviours dependent on learning processes seem to generally trigger frequency dependent selection. That is because the learning algorithm collects more information on the frequent values of the the trait distribution. For the sake of simplicity we modelled a single dimension (badge size). However, this effect could potentially be more important when individuals learn from multidimensional signals.

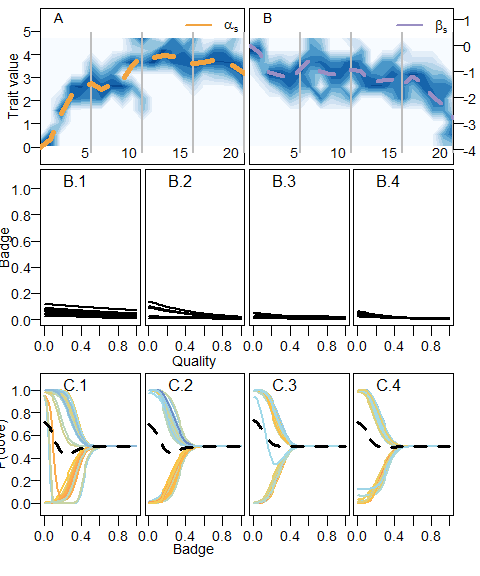
Associate learning is a general cognitive mechanism that allows individuals to associate cues with rewards.

Are bibs real signals of quality? if so are they handicaps or badge of status? Originally Rohwer (1975) proposed the idea that certain morphological traits could be used by animals as status signals or signals of quality. However, from an evolutionary perspective it is unclear how the stability of such communication systems can be maintained. One option is that honesty is maintained by the cost of signal production; that is, when the signal works as a handicap (Botero et al. 2010; Grafen 1990; Johnstone and Norris 1993). Alternatively, if the signal does not carry production costs inversely proportional to quality it may work as a convention. In this later case, honesty is supposedly maintained by the aggressive reaction of receivers when the convention is broken (Enquist, Ghirlanda, and Hurd 2010; Tibbetts and Dale 2004). Our model portraits both kinds of signals, although with some nuances. When we assume the handicap principle, we get the evolution of honest signals mediated by the learned response. When we assume cost-free signals, under certain conditions the learned response mediated the emergence of phenotypic variation. This variation, particularly in the mid-range of the distribution, facilitates the establishment of a convention, in this range individuals respond appropriately to the trait of their peer. The house sparrow (*Passer domesticus*) has been a text-book example of badges of status. According to the prevailing narrative the bib size in males is a signal of the dominance rank and quality. However, a recent meta-analysis called into question this narrative by showing that the effect size of the association between bib size and dominance rank is small and uncertain (Sánchez-Tójar et al. 2018). The simple straight correlation between bib size and dominance rank ( and quality) is expected under the handicap principle, as it is shown here and elsewhere (Botero et al. 2010; Grafen 1990; Johnstone and Norris 1993). But, the diversification process presented here shows that conventions can emerge in a less straightforward way when mediated by learning processes. This nuance could potentially explain seemingly contradictory evidence on the correlation between plumage traits and quality. For example, conventions in our simulations arose only in an intermediate range of the signalling trait, while correlations are often performed across the whole range of variation.

The focus of theoretical and empirical work on communication systems, and particularly badges of status, is often explaining the presence and absence of certain morphological traits within populations. However, the morphological traits hypothesized to play a role in communication systems regularly present interesting patters at the phylogeographic level. One example of this is the leapfrog pattern (Remsen 1984), whereas a certain morphological trait alternates its presence and absence in a set of geographically adjunct populations. Patterns of molecular variation along geographic clines suggest that natural selection is involved in the emergence of these patterns (C. D. Cadena, Cheviron, and Funk 2011; Carlos Daniel Cadena, Klicka, and Ricklefs 2007). However, it is unclear what type of ecological process is behind these selective regimes. Given the wide variation in outcomes found in our model, that depend on both stochasticity and cognitive parameters, we think variation in cognition and learning in particular could provide some explanatory power in this respect. One first step to evaluate the validity of our model, for example, would to evaluate the way individuals in different populations respond to novel traits (Avendaño and Cadena 2021).

We have presented here an evolutionary model where the evolution of a communication system is mediated through the learned responses of receivers. This is a novel way to understand the evolution of communication that integrates a classical cognitive process in learning with evolutionary explanations of communication. This approach contributes to the further integration of proximate and ultimate explanations in behavioral and evolutionary biology.

# Supplementary material



The evolution of cheap signals. Portrait of the evolutionary dynamics of the sender code with snapsots of both sender and receiver codes just as in fig. . The middle panels show changes in the distribution of values for and along evolutionary time. Panels above and below correspond to snapshots of the sender and receiver codes, respectively, generation time of the snapshots are indicated by the grey lines in the middle panels.

Notation of the model parameters and variables

| Symbol | Description |
| --- | --- |
|  | Quality of individual |
|  | Population size |
|  | Standard deviation of the truncated normal distribution from which quality is drawn |
|  | Badge size of individual |
|  | Intercept of the badge size reaction norm for individual |
|  | Slope of the badge size reaction norm for individual |
|  | Mutation rate |
|  | Standard deviation of the normal distribution from which quality is mutations are drawn |
|  | Survival probability of individual |
|  | Intercept of the survival function |
|  | Slope of the survival function |
|  | Value of the contested resource |
|  | Cost of an escalated fight |
|  | Probability that individual wins an escalated fight against individual |
|  | Parameter determining how important is quality defining the probability of wining |
|  | Prediction error |
|  | Speed of learning |
|  | Number of feature centres |
|  | badge size for feature centre |
|  | Reward estimate |
|  | weight of feature centre on the reward estimation |
|  | width of the generalization function |
|  | probability that individual retreats from a fight in an encounter with individual |
|  | weight of feature centre z on the tendency to retreat |
|  | total pay-off of individual |
|  | base line pay-off |

## References

Avendaño, Jorge Enrique, and Carlos Daniel Cadena. 2021. “Territorial Males Do Not Discriminate Between Local and Novel Plumage Phenotypes in a Tropical Songbird Species Complex: Implications for the Role of Social Selection in Trait Evolution.” *Behavioral Ecology and Sociobiology* 75 (2): 37. <https://doi.org/10.1007/s00265-021-02976-8>.

Behrens, Timothy E. J., Laurence T. Hunt, Mark W. Woolrich, and Matthew F. S. Rushworth. 2008. “Associative Learning of Social Value.” *Nature* 456 (7219): 245–49. <https://doi.org/10.1038/nature07538>.

Bond, Alan B., and Alan C. Kamil. 2002. “Visual Predators Select for Crypticity and Polymorphism in Virtual Prey.” *Nature* 415 (6872): 609–13. <https://doi.org/10.1038/415609a>.

Botero, C. A, I. Pen, J. Komdeur, and F. J Weissing. 2010. “The Evolution of Individual Variation in Communication Strategies.” *Evolution* 64 (11): 3123–33.

Bradbury, Jack W., and Sandra L. Vehrencamp. 2011. *Principles of Animal Communication*. 2nd Edition. Sunderland, Mass: Sinauer Associates is an imprint of Oxford University Press.

Cadena, C. D., Z. A. Cheviron, and W. C. Funk. 2011. “Testing the Molecular and Evolutionary Causes of a ‘Leapfrog’ Pattern of Geographical Variation in Coloration.” *Journal of Evolutionary Biology* 24 (2): 402–14. <https://doi.org/10.1111/j.1420-9101.2010.02175.x>.

Cadena, Carlos Daniel, John Klicka, and Robert E. Ricklefs. 2007. “Evolutionary Differentiation in the Neotropical Montane Region: Molecular Phylogenetics and Phylogeography of Buarremon Brush-Finches (Aves, Emberizidae).” *Molecular Phylogenetics and Evolution* 44 (3): 993–1016. <https://doi.org/10.1016/j.ympev.2006.12.012>.

Dridi, Slimane, and Erol Akçay. 2018. “Learning to Cooperate: The Evolution of Social Rewards in Repeated Interactions.” *The American Naturalist* 191 (1): 58–73. <https://doi.org/10.1086/694822>.

Dridi, Slimane, and Laurent Lehmann. 2016. “Environmental Complexity Favors the Evolution of Learning.” *Behavioral Ecology* 27 (3): 842–50. <https://doi.org/10.1093/beheco/arv184>.

Enquist, Magnus, Stefano Ghirlanda, and Peter L. Hurd. 2010. “Signaling.” In *Evolutionary Behavioral Ecology*. Oxford University Press.

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

Fawcett, Tim W., Steven Hamblin, and Luc-Alain Giraldeau. 2013. “Exposing the Behavioral Gambit: The Evolution of Learning and Decision Rules.” *Behavioral Ecology* 24 (1): 2–11. <https://doi.org/10.1093/beheco/ars085>.

Frankenhuis, Willem E., Karthik Panchanathan, and Andrew G. Barto. 2018. “Enriching Behavioral Ecology with Reinforcement Learning Methods.” *Behavioural Processes*, February. <https://doi.org/10.1016/j.beproc.2018.01.008>.

Grafen, Alan. 1990. “Biological Signals as Handicaps.” *Journal of Theoretical Biology* 144 (4): 517–46. <https://doi.org/10.1016/S0022-5193(05)80088-8>.

Guilford, Tim, and Marian Stamp Dawkins. 1991. “Receiver Psychology and the Evolution of Animal Signals.” *Animal Behaviour* 42 (1): 1–14. <https://doi.org/10.1016/S0003-3472(05)80600-1>.

Heyes, Cecilia. 2012. “Simple Minds: A Qualified Defence of Associative Learning.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1603): 2695–2703. <https://doi.org/10.1098/rstb.2012.0217>.

Johnstone, Rufus A. 1997. “Recognition and the Evolution of Distinctive Signatures: When Does It Pay to Reveal Identity?” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 264 (1387): 1547–53. <https://doi.org/10.1098/rspb.1997.0215>.

Johnstone, Rufus A., and Ken Norris. 1993. “Badges of Status and the Cost of Aggression.” *Behavioral Ecology and Sociobiology* 32 (2): 127–34. <https://doi.org/10.1007/BF00164045>.

Kamil, Alan C. 1983. “Optimal Foraging Theory and the Psychology of Learning.” *Integrative and Comparative Biology* 23 (2): 291–302. <https://doi.org/10.1093/icb/23.2.291>.

Leimar, Olof, and Redouan Bshary. 2022a. “Reproductive Skew, Fighting Costs and Winner–Loser Effects in Social Dominance Evolution.” *Journal of Animal Ecology* n/a (n/a). <https://doi.org/10.1111/1365-2656.13691>.

———. 2022b. “Effects of Local Versus Global Competition on Reproductive Skew and Sex Differences in Social Dominance Behaviour.” *Proceedings of the Royal Society B: Biological Sciences* 289 (1987): 20222081. <https://doi.org/10.1098/rspb.2022.2081>.

Leimar, Olof, and John M. McNamara. 2019. “Learning Leads to Bounded Rationality and the Evolution of Cognitive Bias in Public Goods Games.” *Scientific Reports* 9 (1): 1–9. <https://doi.org/10.1038/s41598-019-52781-7>.

Macphail, Euan M. 1982. *Brain and Intelligence in Vertebrates*. Clarendon Press.

Maynard-Smith, John. 1982. *Evolution and the Theory of Games*. 1 edition. Cambridge ; New York: Cambridge University Press.

McAuliffe, K., and A. Thornton. 2015. “The Psychology of Cooperation in Animals: An Ecological Approach.” *Journal of Zoology* 295 (1): 23–35. <https://doi.org/10.1111/jzo.12204>.

Møller, Anders Pape. 1988. “Female Choice Selects for Male Sexual Tail Ornaments in the Monogamous Swallow.” *Nature* 332 (6165): 640–42. <https://doi.org/10.1038/332640a0>.

Parker, G. A. 1974. “Assessment Strategy and the Evolution of Fighting Behaviour.” *Journal of Theoretical Biology* 47 (1): 223–43. <https://doi.org/10.1016/0022-5193(74)90111-8>.

Quiñones, Andrés E., G. Sander van Doorn, Ido Pen, Franz J. Weissing, and Michael Taborsky. 2016. “Negotiation and Appeasement Can Be More Effective Drivers of Sociality Than Kin Selection.” *Phil. Trans. R. Soc. B* 371 (1687): 20150089. <https://doi.org/10.1098/rstb.2015.0089>.

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

Remsen, J. V. 1984. “High Incidence of "Leapfrog" Pattern of Geographic Variation in Andean Birds: Implications for the Speciation Process.” *Science* 224 (4645): 171–73. <https://doi.org/10.1126/science.224.4645.171>.

Rohwer, Sievert. 1975. “The Social Significance of Avian Winter Plumage Variability.” *Evolution* 29 (4): 593–610. <https://doi.org/10.2307/2407071>.

Sánchez-Tójar, Alfredo, Shinichi Nakagawa, Moisès Sánchez-Fortún, Dominic A Martin, Sukanya Ramani, Antje Girndt, Veronika Bókony, et al. 2018. “Meta-Analysis Challenges a Textbook Example of Status Signalling and Demonstrates Publication Bias.” Edited by Diethard Tautz and Tim Parker. *eLife* 7 (November): e37385. <https://doi.org/10.7554/eLife.37385>.

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

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

Tibbetts, Elizabeth A., and James Dale. 2004. “A Socially Enforced Signal of Quality in a Paper Wasp.” *Nature* 432 (7014): 218–22. <https://doi.org/10.1038/nature02949>.

Zahavi, Amotz. 1975. “Mate Selection—-A Selection for a Handicap.” *Journal of Theoretical Biology* 53 (1): 205–14. <https://doi.org/10.1016/0022-5193(75)90111-3>.

1. Laboratorio de Biología Evolutiva de Vertebrados, Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá, Colombia [↑](#footnote-ref-1)
2. Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland [↑](#footnote-ref-2)