Andrés E. Quiñones[[1]](#footnote-1)

C. Daniel Cadena\*

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

January 27, 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 conspecific individuals 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 engaging in a social interaction might vary depending on its 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 such 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 individuals may benefit from broadcasting accurate information, whereas others may benefit from concealing it. Take, for instance, an interaction between two individuals where one can help the other. Given some costs and benefits, a potential donor would be interested in helping related individuals. Therefore, for a relative of the donor, broadcasting kinship would be advantageous, whereas concealing the lack of kinship would be better for an unrelated individual. Similar scenarios may apply to a variety of aspects of social life such as finding mates, feeding offspring, or engaging in dominance relationships or aggressive interactions (Bradbury and Vehrencamp 2011; Møller 1988; Tibbetts and Dale 2004).

In dominance and aggressive interactions, a crucial piece of information to guide individual actions is the fighting ability of interacting partners, which is sometimes referred to as Resource Holding Potential (RHP; Parker 1974), or simply as quality. Responsiveness to the quality of an opponent is central to communication systems such as those in which individuals exhibit badges of status, where a signal conveys quality (Johnstone and Norris 1993; Rohwer 1975). Such signals are often arbitrary, meaning they are 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 is 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 are cognitive aspects of the receiver module. Theoretical models often assume that communication systems relying on badges of status 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 contests. Reaction norms allow individuals to respond to the available information without a large cognitive burden. That is at least in contrast to other information processing mechanisms like individual recognition. With individual recognition, individuals associate cues of their peers to their quality. Because these associations must be learnt throughout the life of individuals, they have the usual cognitive requirements of associative learning processes. 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 with attributes of the badge and would be reinforced by every interaction. Therefore, in principle, learning in such scenarios would be faster than in systems involving individual recognition in which the association between signals and their meaning varies depending on the interacting partners. In any case, learning may 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 taxonomically widespread cognitive mechanism that allows individuals to associate rewards with environmental stimuli and thus behave adaptively (Staddon 2016) (Heyes 2012; Macphail 1982; Staddon 2016; Behrens et al. 2008). Theory has shown that natural selection favours such associations in complex environments where conditions are difficult to predict (Dridi and Lehmann 2016). Associative learning is a flexible cognitive mechanism whose underpinnings show interspecific variation (Enquist, Lind, and Ghirlanda 2016; Quiñones et al. 2019), which suggests it is amenableto evolution by natural selection. 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 the evolution of behaviour. 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).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 resource, depending on a quantitative morphological trait (i.e. a badge indicating quality) they perceive in their opponents. Over evolutionary time, the size of badge evolves, and whether it depends on their 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 indicating individual 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 resource holding potential, whereas an individual with quality has the highest. As they develop, individuals produce a phenotypic signal (badge), the size of which depends on their 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 to take values 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. The survival probability of any given individual is given by , where and are parameter values determining the shape of the survival function. Importantly, if the 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. Given that individuals in real populations typically do not interact at random nor with all individuals in the population (Kurvers et al. 2014), we use the population vector to define an interaction neighbourhood for the focal individual where the size of the neighbourhood () is the number of positions around the focal individual from which the partner is drawn. So, if interactions are global; as gets smaller, interactions are more local. In each interaction, individuals 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, then they split up the resource in half; if both decide to fight, then 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 it and its interacting partner:

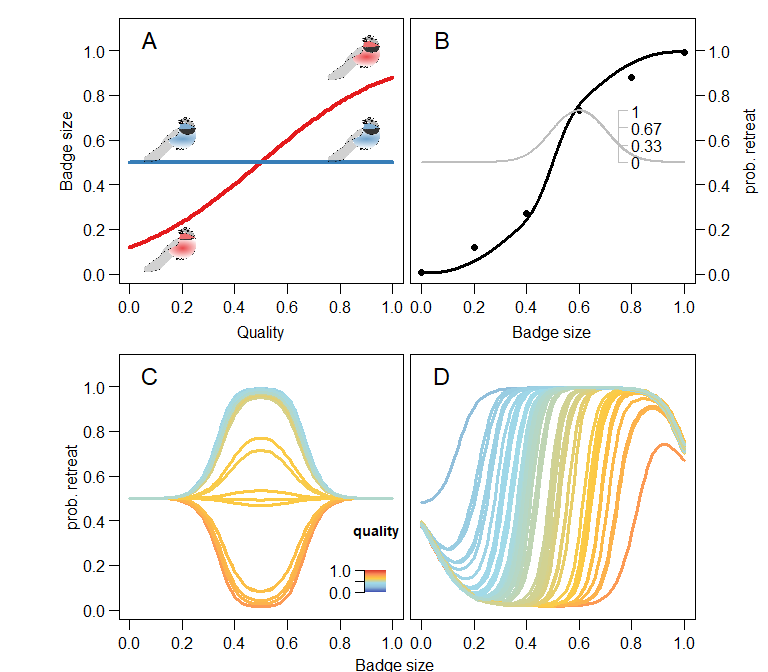
, where is a parameter defining how strong the quality difference determines the winning probability, and and denote the position in the population vector of the focal and its interacting partner, respectively.

The decision of whether to escalate a fight against an interacting partner can depend 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 each 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, individuals 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 this leads to an increase in the reward estimation, then 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/attacking, 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]). This feature centres are the same for all individuals and stored in vector . Each of these feature 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, weighted by the response triggered by the feature (Fig. B, dots represent the feature weights) The response of each feature center 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.

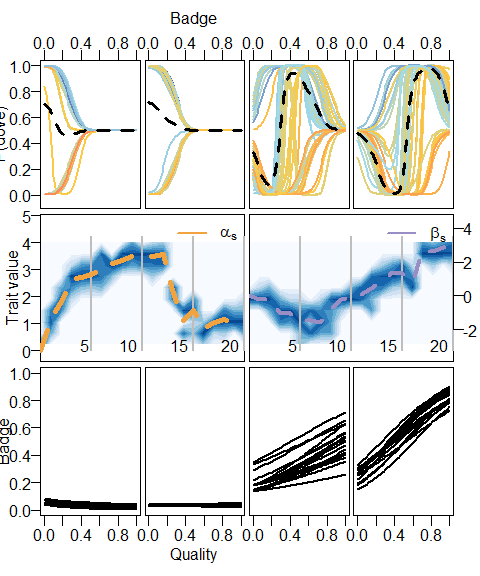
After all their interactions, individuals in the population reproduce proportionally to their total pay-off , which is a sum of the baseline pay-off () and all the pay-offs obtained throughout their life. Thus, the combination of natural selection and genetic drift changes the distribution of values 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 and 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. The simulations reveal that when learners face uninformative signals, they modify 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; whereas individuals of lower quality (blue tones) mostly retreat from confrontations. Thus, learning splits the population of receivers into the 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 decision to retreat or attack depends on the quality of the receiver. As expected, the higher the quality of the receiver, the larger the badge size of the opponent that triggers a retreat.

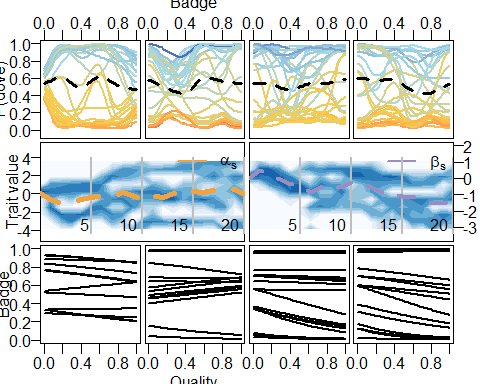


Evolution of the badge size as a handicap mediated by learning. The middle panels show the evolutionary dynamics of the sender code. On the left, changes in the distribution of values for the intercept of the reaction norm (); on the right changes in the distribution of the slope (). Dashed lines in both panels show the mean of the distribution. Grey lines show the generation time corresponding to the panels above and below showing the sender and receiver code. In the panels above, the reaction norms correspond to individuals in the population after the interaction round. Color scale indicates quality just as in Fig. . In the panels below, colors represent a cluster clasification performed by the k-means algorithm

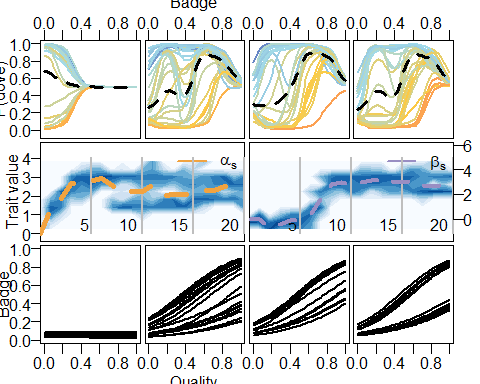
When we allow the sender code to evolve ( and change subject to natural selection and genetic drift) and the badge size works as a handicap (i.e. the cost of the badge is inversely proportional to the quality of the individual), the sender code can evolve to produce an honest signal Fig ). The evolution 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. , X) appear to involve a set of steps. First, the intercept of the reaction norm () describing the relationship between individual quality and badge size evolves to higher values, reducing the average badge size in the populations. This is because reaction norms segregating in the population are flat at this point; consequently, the receiver code does not respond to the badge size, and larger badges are costlier and do not trigger lower attack probabilities. During the first generations, the slope of the sender reaction norm () remains close to the initial value of zero. At around generation 4000, the slope evolves toward positive values, which implies that larger badges correlate with higher individual quality. Receivers then learn to react to such correlation, reducing the probability of attack towards individuals with larger badges. Hence, natural selection favours larger values of , eventually leading to an evolutionary equilibrium in which badge size is an honest signal of quality mediated by the learned responses of receivers.

The evolutionary trajectory portrayed in figure is not the only possible outcome. 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 individuals (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 these cases reaction norms are flat, so the badge does not provide any information about quality.

The amount of information that agents are able to collect through learning throughout their life time can strongly change the outcome of evolutionary dynamics. In the simulations presented in figure and , individuals learned with high learning rate () and interacted repeatedly along their life (on average 2000 interactions). When we reduced the number of interactions that individuals have over their lifetime to 300 on average, we saw a drastic increase in the phenotypic and genetic variation present in the population. Fig. shows changes in the distribution of the sender reaction norm along evolutionary time assuming the badge is cost-free. 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 value in the population shows three peaks. In the case of , the peaks in the distribution are not so clear-cut, but it is clear that there is an increase in the variance of the distribution. These changes in the distribution of the parameters of the sender reaction norm imply that during the second half of the simulation individuals can be classified into three distinct types. Two types express a flat reaction norm, meaning that their badge size is not informative of their quality, whereas the third type shows intermediate badge sizes determined by the quality of the individual (Fig. ). 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 larger badges (Fig. ).

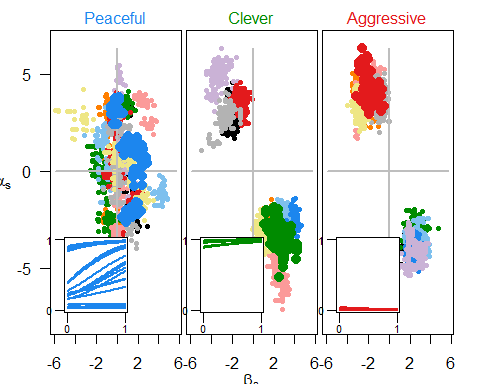


A limited number of interactions on the amount of variation has an effect on the evolving parameters when the signal follows the handicap principle. In fig , we show simulations where individuals have on average 300 interactions in their life and the cost of the signal is proportional to the quality, 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 muddle by the fact that there are two clusters of values for the slope of the reaction norm () in the population. Thus, there are two types of reaction norms. One of the types has 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 can acquire through learning allows the coexistence of different communication strategies within a population.



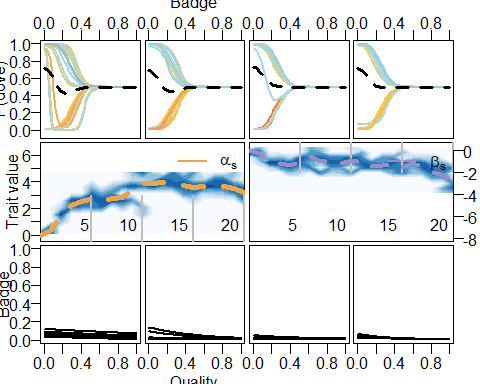
The evolution of diverse costly signals. Portrait of the evolutionary dynamics of the sender code with snapsots of both sender and receiver codes just as in fig. . The middle panels show changes in the distribution of values for and along evolutionary time. Panels above and below correspond to snapshots 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 assess whether such 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) low (aggressive) probability of escalating a fight and 3) corresponding to the evolutionarily stable strategy 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, corresponding to peaceful naive individuals, is the only one where the  
distribution of values is split in different clusters. That is, most 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 differences among replicates. 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, individuals 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. Creo que una cosa clave es trabajar en las leyendas de forma que sean explicitas en indicar cuáles son los puntos importantes que revelan las diguras desde el punto de vista biológico y de las preguntas que abordamos.

# 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

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>.

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.

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

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.” *Phil. Trans. R. Soc. B* 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.” *Behav Ecol Sociobiol* 32 (2): 127–34. <https://doi.org/10.1007/BF00164045>.

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

Kurvers, Ralf H. J. M., Jens Krause, Darren P. Croft, Alexander D. M. Wilson, and Max Wolf. 2014. “The Evolutionary and Ecological Consequences of Animal Social Networks: Emerging Issues.” *Trends in Ecology & Evolution* 29 (6): 326–35. <https://doi.org/10.1016/j.tree.2014.04.002>.

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.” *J Zool* 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>.

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

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)