



Original Article

Bayes' rule and bias roles in the evolution of decision making

Sergio Castellano^a

^aDepartment of Life Sciences and Systems Biology, Università di Torino, Via Accademia Albertina, 13, 10123 Torino, Italy

Received 11 September 2014; accepted 21 September 2014; Advance Access publication 4 November 2014.

Behavior is often plastic and the study of the functional basis of behavior should provide insights into the adaptive role of the mechanisms responsible for behavioral flexibility. Cognitive biases provide a window into the psychological mechanisms of behavior and a functional theory of behavioral mechanisms should be able to explain also the evolutionary significance of cognitive biases, that is, whether they should be seen as a solution to environmental problems or more as a by-product of adaptive rules. But how should such a theory be developed? In behavioral ecology, the “heuristics” approach has been prevailing. It proposes the rule first and then it investigates how the rule should be optimally modified to solve the problems posed by the environment, cognitive biases being a possible solution. Here, I explore an alternative approach, which derives rules from formal models of optimality. By focusing on mate choice, I present an optimal Bayesian decision-making model, based on the computation of the cumulative sums of the log-likelihood ratios that the quality of a prospective mate is either higher or lower than average. When uncertainty is high, log-likelihood ratios can be approximated to a linear function of mate quality; specifically, they may be expressed as deviations of the perceived quality from the population mean, weighted by the perceived assessment accuracy. I use this model to illustrate the different origins and, possibly, the different functional role of decision-making biases. More generally, I propose the model as an attempt to develop an evolutionary theory of behavioral mechanisms.

Key words: behavioral plasticity, cognitive bias, decision making, heuristic, mate choice.

INTRODUCTION

Behavioral ecology has been traditionally interested in the functional explanation of behavior and, to this purpose, it has developed a coherent body of theory based on normative models of optimality (Parker and Smith 1990). By assuming behaviors as the target of selection, these models describe which behaviors natural selection is expected to promote. However, behaviors are highly flexible phenotypic traits and selection is expected to act on the rules that govern behavioral flexibility. For this reason, in the last few years, some researchers have been arguing that behavioral ecologists should shift their focus from the functional role of single behaviors to the functional role of the mechanisms underlying behavioral flexibility, such as learning and decision-making mechanisms (McNamara and Houston 2009). This new approach is still underexplored (Fawcett et al. 2013). One possible reason for this is the lack of a coherent theory that could guide the empirical research. Are traditional models of optimality, based on the principles of economical rationality, still a useful tool or should we look for more effective substitutes?

When we focus on the mechanisms of decision making, it appears evident that the patterns of choice are often far from being rational. Cognitive psychologists have described a large number of biases in humans and they have interpreted such departures from rational behavior as evidence that our decisions under uncertainty often rest on a limited number of simple rules (heuristics), rather than on extensive algorithm processing (Gilovitch et al. 2002). A cognitive bias is “a consistent deviation from an accurate perception of judgment of the world” (Fawcett et al. 2014, p. 153). Although cognitive psychologists agree about the important role of biases in unraveling the psychological mechanisms of human reasoning, they often disagree on the functional significance to assign to cognitive biases. Some cognitive psychologists view biases as systematic errors in rational reasoning, thus as constraints that the underlying heuristics impose on decisions (Tversky and Kahneman 1974). Others tend to assign a more positive role to cognitive biases. They see biases as the expression of a new form of rationality, the ecological rationality, which defines an optimal decision as the optimal use of information made available by the environment (Goldstein and Gigerenzer 2002). According to this view, the theoretical research, rather than focusing on biologically implausible models of optimal decisions, such as Bayesian models, should look for fast-and-frugal heuristics, which may not provide the optimal solution in all possible contexts,

Address correspondence to S. Castellano. E-mail: sergio.castellano@unito.it.

but which do perform well in the ecologically relevant conditions experienced by decision makers (Gigerenzer et al. 2002).

Irrational biases are not exclusive to our species. They have been described for other organisms as well, from slime moulds (Latty and Beekman 2011) to bees (Shafir et al. 2002; Bateson et al. 2011), from birds (Bateson et al. 2003; Waite 2008) to mammals (Bentosela et al. 2009), suggesting the intriguing hypothesis that similar biases may result from functionally similar decision rules. For this reason, in the last few years, cognitive biases have received the increasing attention of behavioral ecologists. The empirical research has found in cognitive biases a tool to investigate the psychological mechanisms of behavior (Bateson and Healy 2005), whereas the theoretical research has focused on their functional role, testing the hypothesis of ecological rationality (review in Fawcett et al. 2014).

Two approaches have guided the theoretical research on cognitive bias. The first is best represented by the work of Johnson and Fowler (2011) on the functional role of overconfidence in antagonistic interactions. In their game-theoretical model, both the decisional rule and the type of available information are given as assumptions. The model shows that overconfidence is an evolutionarily stable strategy, because it is what allows decision makers to optimally adapt their rule to the characteristics of their environment (Johnson et al. 2013). This approach closely resembles the “fast-and-frugal heuristics” research program in cognitive psychology in that it proposes the rule first and then it shows how the rule can be optimally expressed under different environmental conditions. According to this view, the rule is the constraint and the cognitive bias is the ecologically rational trait, the target of selection.

The second approach to the theoretical study of cognitive biases is that advocated by Marshall et al. (2013), who argue that an evolutionary theory of cognitive bias requires a close integration of function and mechanism. According to this approach, optimal rules should be “a-priori” derived from formal optimality models, they should not be a given, but the outcome of adaptive evolution. In non-game-theoretic decision making, optimal rules are derived from unbiased estimates of Bayesian posterior probabilities. Since these estimates are computationally highly demanding, natural selection is thought to favor decision rules that approximate Bayesian computation. Theory should thus look for algorithms that could effectively, though imperfectly, implement optimal decisional rules. In the present work, I will pursue this line of theoretical research. In the next section, I present the rationale of this approach. Then, I present a model of optimal decision making based on Bayesian computation of posterior probabilities and derive from it a plausible algorithm that animals could use to make decisions. Finally, I use this model to provide insights into the functional role of cognitive bias and of the mechanisms underlying behavioral flexibility.

The decision-making mechanism

“Decision is a deliberative process that results in the commitment to a categorical proposition” (Gold and Shadlen 2007, p. 536). In the case of dichotomous choices, from a computational point of view, a decision is the mapping of a continuous variable (the decision variable) into a binary variable (the adopted decision) by means of a decision rule. In this sense, it is equivalent to a classification process, which assigns an object to either one of 2 mutually exclusive groups by using a discriminatory criterion. The decision variable is the amount of evidence that the object belongs to the group. The decision rule is the threshold value, that is, the minimum amount of evidence to belong to that group.

Animals can make decisions because their nervous system can carry out this kind of computation. As Marr (1982) clearly

explained some decades ago, if we are to understand how such a device works, we need to analyze the problem, at least, at 3 levels: 1) at the abstract-computational level, by analyzing the functional role of the device (i.e., we should ask what it does and why); 2) at the representation-algorithm level, by asking how information is represented and which are the algorithms used to process it; and 3) at the proximate level by asking how the nervous system physically represents information and implements algorithms.

If we are to understand why the information-processing device of decision making is biased, we need to investigate the problem at both the first and second levels described by Marr. In Figure 1, I show the rationale of this approach. At the abstract level of computational theory, the model is concerned only with the principles of optimality, it is “fully” rational and no biases are expected to arise. At the algorithmic-representation level, 2 types of computational bias may arise. I call the first type of bias the “representation” bias, because it is due to systematic errors in the representation of computational objects. For example, small brains with a low computational capability may make linear approximations of nonlinear operators, providing a biased representation of the variables involved in computation (Lange and Dukas 2009). I call the second type of algorithm bias the “parameterization” bias, because it arises as the effect of systematic errors in the parameters used by the device to make the algorithm work. For example, Johnson and Fowler (2011) suggest that overconfidence is the effect of a parameterization bias in the evaluation of the decision maker’s quality. Both types of algorithm biases are, by

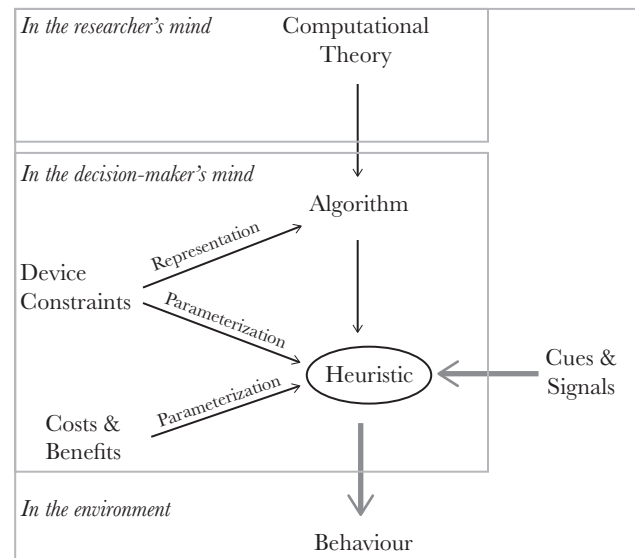


Figure 1

The rationale of the computational theory approach to the modeling of decision making mechanisms. The computational theory is seen as a tool to derive hypotheses about the mechanism (algorithm) of decision making. Whether the computational theory shows optimal solutions to computational problems, the actual solution (algorithm) implemented in the device may only approximate optimality because of the device’s constraint. Constraints act at 2 levels: at the representation level, they arise from the way information is represented and they affect both the accuracy and the precision of the evaluation process; at the parameterization level, the constraints directly act on the parameters that make the algorithm work. For example, previous experience may constrain the parameters to a range of values that make the algorithm work suboptimally if the choice context has changed. Decision makers can adjust the rule strategically, depending on the costs and benefits imposed to them by the environment. Adjustments are supposed to involve the parameters rather than the structure of the decision mechanism.

definition, “cognitive biases,” though their role in the evolution of decision making may be very different.

To present and defend my thesis, first I illustrate a model of decision making at both the abstract-computational level and the representation-algorithm level. Then, I will use this model to investigate the different effects of computational and representation biases on decision making and their different evolutionary roles.

The model

Step 1: the optimal Bayesian choice model

I use a mate-choice scenario to illustrate how an optimal cognitive machinery for decision making should work. At this stage, the question I face is not about the effects of costs and benefits of mate choice on optimal mating decisions, but about the computational mechanisms that allow an animal to choose optimally, once the costs and the benefits of choice are known.

The model makes the following assumptions: 1) decision makers (hereafter females) evaluate prospective mates (hereafter males) one at a time and continue to assess until either accepting or rejecting them (sequential search model); 2) females use cues to assess male quality and males have no “strategic” control over what females perceive of their quality; 3) male “true” quality is normally distributed with mean θ , variance σ_{pop}^2 , probability density function $g(q)$, and probability cumulative distribution function $G(q)$; 4) female assessment of male quality is highly inaccurate and the perceived quality of a prospective mate of “true” quality q is $Q = q + \epsilon$, where ϵ is a normally distributed random variable with 0 mean and variance σ_{err}^2 (Figure 2a).

Depending on the costs and benefits of choice (see below), there exists an optimal quality threshold, q^* , so that only mates with quality $q \geq q^*$ should be chosen as mates. Since male-quality evaluation is prone to error ($\sigma_{err}^2 \gg 0$), the optimal decision should be based on the computation of $P(H|Q)$, the posterior probability that the hypothesis H ($H:q > q^*$) is true given the perceived quality Q of the prospective mate. The Bayes theorem explains how this posterior probability should be computed:

$$P(H|Q) = \frac{P(Q|H)P(H)}{P(Q)} \tag{1}$$

where $P(Q|H)$ is the conditional probability of observing Q when H is true, $P(Q)$ is the probability of observing Q independent of H being true or false, and $P(H)$ is the prior probability that H is true (i.e., $P(H) = 1 - G(q)$). If we define H' as the alternative hypothesis of H (i.e., $H':q < q^*$), so that $P(H') = 1 - P(H)$, we may use Equation 1 to express the posterior probability in terms of log-odds ratios:

$$\text{Log} \left(\frac{P(H|Q)}{P(H'|Q)} \right) = \text{Log} \left(\frac{P(Q|H)}{P(Q|H')} \right) + \text{Log} \left(\frac{P(H)}{P(H')} \right) \tag{2}$$

Denoting the log odds-ratio of the posterior probability by V (the decision variable), the log odds-ratio of the prior probability by V_0 , and the logarithm of the likelihood ratio by $I(Q, H)$ (the amount of acquired evidence supporting the hypothesis H), the Equation 2 may be rewritten as:

$$V = I(Q, H) + V_0 \tag{3}$$

A decision consists of the mapping of V into a binary variable D by applying a decision rule $d(V)$. For example, the following decision rule,

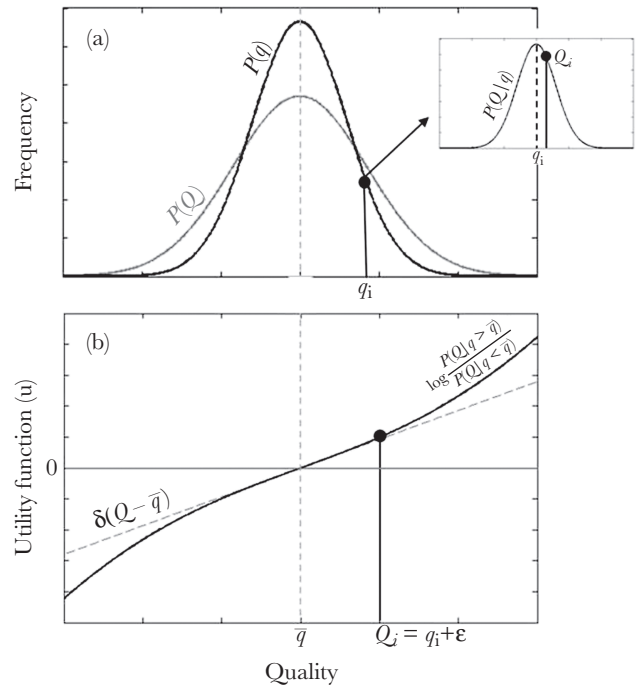


Figure 2

The Approximate Bayesian Decision Making model. (a) The quality of males in the population is assumed to be normally distributed and $P(q)$ is its probability density function, which has mean θ and variance σ_{pop}^2 . Quality assessment, however, is highly uncertain and a male of quality q_i can be perceived of quality Q_i with probability $P(Q_i|q_i)$. Conditional probabilities are assumed to be normally distributed with mean q_i and variance σ_{err}^2 . $P(Q)$ is the normal probability density function of males’ perceived qualities, it has mean θ and variance $\sigma_{pop}^2 + \sigma_{err}^2$. (b) A perceived quality Q_i is converted in the perceived utility u_i by the linear approximation of the log-odds ratio function of Q . The solid line represents the log-likelihood ratio of $P(Q_i|H)$ and $P(Q_i|H')$ (i.e., the OBC model), whereas the dashed line is the linear approximation of the log-likelihood ratio (i.e., the ABC model). The regression coefficient (δ) of the linear approximation depends only on σ_{pop}^2 and σ_{err}^2 and increases with the increasing assessment accuracy (see Appendix 1).

$$d(V) = \begin{cases} \text{if } V > 0 \rightarrow D = 1 \text{ (commit to } H) \\ \text{if } V < 0 \rightarrow D = 1 \text{ (commit to } H') \end{cases} \tag{4}$$

will force females to choose a prospective mate when V is positive and to reject him when V is negative. The model described by Equations 3 and 4 is inherently static, because it does not include time in the decision process. Implicitly, it assumes that the decision time and, thus, the amount of evidence needed to make a decision about a prospective mate is fixed and independent of his quality (Castellano 2009). To include time in the decision process, I now assume that females obtain a sequence of noisy observations (Q_1, Q_2, Q_3, Q_4). The decision variable V_t after t evaluations (of the same male), is the sum (or, equivalently, the arithmetic mean) of all the single evaluations obtained (Figure 3).

$$V(t) = V(t - 1) + I(Q(t), H), \quad V(0) = V_0 \tag{5}$$

Under this assumption, the decision process is equivalent to the Sequential Probability Ratio test (SPRT) (Wald 1945), which is an “optimal” decision-making procedure, in that, for a given error rate, it minimizes the average number of samples required to choose between the 2 alternative hypotheses (Wald and Wolfowitz 1948; Bogacz et al. 2006). Let β be the probability of choosing an

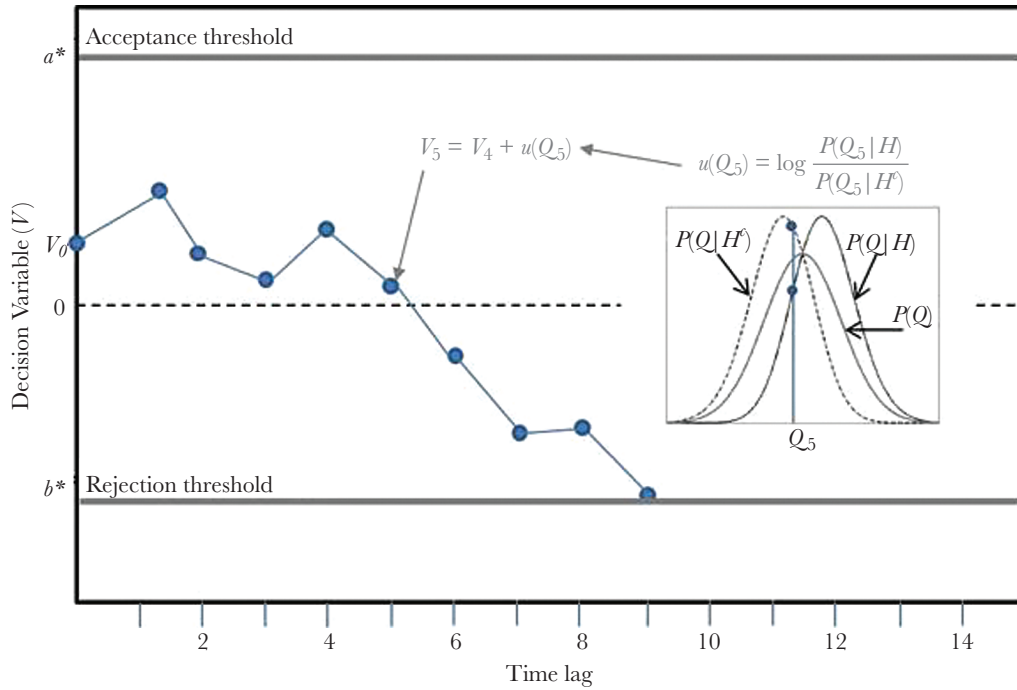


Figure 3

Graphical representation of an optimal decision mechanism between 2 mutually exclusive hypotheses. At discrete-time intervals, the decisional device receives statistically independent pieces of noisy information about the tested hypotheses. It converts the information into conditional probabilities ($P(Q|H)$ and $P(Q|H')$), computes their log ratio (the perceived utility of information), and stores the processed information into an accumulator (the decision variable, V). The behavioral response is given as soon as the accumulator reaches either one of the 2 decision thresholds (the acceptance threshold a^* or the rejection threshold b^*). Notice that in this example, the starting value of the decision variable (i.e., V_0) is greater than 0 (dashed line), because the prior probability of H is assumed to be greater than 0.5.

inappropriate mate (i.e., a “false positive” or Type II error rate) and α the probability of rejecting an appropriate mate (i.e., “false negative” or Type I error rate). Depending on their costs, there will be 2 optimal threshold values, α^* and β^* , defining the 2 optimal decision thresholds, $z_1^* = \log \frac{1-\beta^*}{\alpha^*}$ and $z_2^* = \log \frac{\beta^*}{1-\alpha^*}$, and the optimal decision rule $d^*(V)$, so that:

$$d^*(V) = \begin{cases} \text{if } V > z_1^* \rightarrow D = 1(\text{commit to } H) \\ \text{if } z_2^* \leq V \leq z_1^* \rightarrow \text{continue monitoring} \\ \text{if } V > z_2^* \rightarrow D = 1(\text{commit to } H') \end{cases} \quad (6)$$

To reduce error probability (i.e., to improve accuracy), the bounds must be moved further from 0. This, of course, would increase the number of samples (and time) needed to stop the process and, thus, it would increase the costs of making decisions.

Step 2: the approximate Bayesian choice model

The optimal Bayesian choice (OBC) model presented above makes 2 unlikely assumptions: 1) an accurate assessment of inaccuracy (the decision maker must accurately know how inaccurate is its evaluation) and 2) the absence of constraints on the computational power of decision makers. Although the first of these assumptions actually poses no problems to the plausibility of the model (it simply explains why decisions are sometime wrong and almost always suboptimal), the second assumption does pose serious problems because it contests that natural selection could have ever favored the evolution of such a computationally complex mechanism of decision making. Indeed, the most serious concerns raised by the

detractors of the Bayesian approach in the study of the proximate mechanisms of decision making is the difficulty of implementing Bayesian computations in nervous systems (Bowers and Davis 2012, but see McNamara and Houston 1980 and Griffiths et al. 2012).

The representation-algorithm level of analysis must address the question of how the cognitive machinery of decision making creates an internal representation of the main quantities involved in the computation, namely the log-likelihood ratios, $I(Q, H)$. The approximate Bayesian choice (ABC) model is based on the assumption that females do not “know” either the probability density function of male quality ($g(q)$) or the likelihood functions ($P(Q|H)$ and $P(Q|H')$), but that they do know the population mean (θ) and variance (σ_{pop}^2) and the assessment error (σ_{err}^2) of male quality. In the ABC model, females are assumed to represent log-likelihood ratios by means of a linear function $U = u(Q)$, which I name the “utility function” and which is a good linear approximation of $I(Q|H)$ when the hypothesis being tested is that the prospective mate has larger-than-average quality (i.e., $H:q > \theta$). Indeed, under this hypothesis, it may be shown (see Appendix 1) that the 2 likelihood functions, $P(Q|H)$ and $P(Q|H')$, have the same standard deviation, $\sigma_{q|H} = \sigma_{err}^2 + 0.363 \cdot \sigma_{pop}^2$, and mean $(\theta + m)$ and $(\theta - m)$, respectively (where $m = 0.798 \cdot \sigma_{pop}$). When the assessment error (σ_{err}^2) has magnitude similar to the variance in male quality (σ_{pop}^2), then $P(Q|H)$ and $P(Q|H')$ can be approximated to 2 normal functions and $u(Q)$ becomes a linear function of Q (Figure 2b):

$$U(Q) = \delta (Q - \theta) \quad (7)$$

where the regression coefficient, $\delta = \frac{2m}{\sigma_{q|H}^2}$, describes the assessment accuracy. When σ_{err} is much larger than σ_{pop} , then δ tends

to 0 and Q does not reduce uncertainty in mate-quality assessment. When σ_{err} is much smaller than σ_{pop} , δ tends to increase to a maximum of $4.391 \cdot \sigma_{pop}$, and the “usefulness” of the information acquired increases as well. The perceived utility can be used to compute V , the decision variable, and implemented in the random-walk process that describes the dynamic of decision making.

$$V(t) = V(t - 1) + \delta(Q(t) - \theta), \quad V(0) = \frac{\log(P(H))}{P(H^c)} \quad (8)$$

Due to the linear approximation, the ABC model underestimates the log-odds ratios of higher-than-average values and overestimates those of lower-than-average values (Figure 3b). A cognitive machinery that computes the perceived utilities of a stimulus in terms of weighted deviations from its expected value will thus provide a biased representation of the acquired information.

Optimal decision strategy: the effects of parameterization biases on decision making

Now, I address the question of how costs and benefits affect optimal decisions. I assume that costs are directly proportional to the time spent by females to choose a partner and that benefits depend only on the quality q of the chosen male. I define: 1) S the decision strategy adopted by a female; 2) $\Phi(q; S)$ the probability that a female adopting a decisional strategy S will mate with a male of quality q ; 3) $E[\Phi(S)]$ the mean mating probability; 4) $\Psi(q; S)$ the average time spent by that female either to accept or reject a prospective mate of quality q ; and 5) $E[\Psi(S)]$ the average evaluation time of a prospective mate. The benefits of this strategy are the mate expected (mean) quality:

$$B(S) = \frac{\int \Phi(q; S) \cdot g(q) \cdot q \cdot dq}{E[\Phi(S)]} \quad (9)$$

whereas the cost of the strategy is the average time spent to find an appropriate mate:

$$C(S) = \frac{\Psi_0 + E[\Psi(S)]}{E[\Phi(S)]} \quad (10)$$

Ψ_0 is the average searching time and it is assumed to be constant and independent of male quality, whereas $1/E[\Phi(S)]$ is the mean number of males that a female is expected to evaluate. Notice that I do not consider any explicit cost for acquiring information and, thus, I implicitly assume, that the computation costs of $I(Q)$ and $u(Q)$ are equivalent.

Finally, I assume that natural selection maximizes the difference between the expected benefits and the expected costs, that is:

$$W(S) = B(S) - kC(S) \quad (11)$$

where k is a calibration coefficient describing the weighting of searching and decision cost relative to reward.

The decision-making strategy in the OBC and the ABC models

In the OBC model, a decision strategy S_{OBC} is defined by 2 variables: \tilde{q} , the quality threshold, and z , the decision-variable

threshold. The optimal S^*_{OBC} is thus defined by the pair of values $\{q^*, z^*\}$ that maximizes $W(S)$, so that $W(S^*_{OBC}\{q^*; z^*\}) \geq W(S_{OBC}\{\tilde{q}; z\})$, for any q and z . Notice that, in the OBC model, V_0 depends on \tilde{q} (i.e., $V_0(\tilde{q}) = \text{Log} \frac{1-G(\tilde{q})}{G(\tilde{q})}$). So that the optimal V^*_0 is $V^*_0 = V_0(\tilde{q}^*)$.

In the ABC model, the mean population quality θ , the initial value of the decision accuracy variable $V_0 = 0$ (i.e., the prior log ratio), and the assessment accuracy δ are parameters rather than variables of the model. To investigate the functional role of parameterization biases on optimal decisions, I assume that females can introduce systematic errors in the perception of these parameters. Thus, I define: 1) \tilde{q} , the perceived mean population quality and $b_0 = (q - \theta)$ its associated bias (or systematic error); 2) d , the perceived accuracy, and $b_\delta = (d - \delta)$ its bias; and 3) $b_{V_0} = V_0$ the bias of the initial condition of the evaluation process. The decision strategy S_{ABC} is defined by these 3 variables and by z , the decision-variable threshold. A decision strategy $S^*_{ABC}\{\tilde{q}^*; d^*; V_0^*; z^*\}$ is optimal if $W(S^*_{ABC}) \geq W(S)$ for any q, d, V_0 , and z . In Appendix 2, I show that a bias in the perceived accuracy (i.e., $b_\delta \neq 0$) has no effects on the optimal decision, because for any strategy S_{ABC}

$\{\tilde{q}; d; V_0; z\}$, there is an infinite number of alternative strategies $S^*_{ABC}\left\{\tilde{q}; (d + b_\delta); \left(V_0 \frac{d + b_\delta}{d}\right); \left(z \frac{d + b_\delta}{d}\right)\right\}$ for which $\Phi(q; S) =$

$\Phi(q; S^*)$ and $\Psi(q; S) = \Psi(q; S^*)$. For this reason, in the following sections, I will describe the ABC decision strategy only in terms of q, V_0 , and z .

Choice probability and decision time

As indicated above, optimal decision strategy is the one that maximizes the weighted difference between the expected mate quality, which depends on $\Phi(q; S)$, and the average decision time, which depends on $\Psi(q; S)$. Unfortunately, the random-walk process described by Equations 5 and 8 does not yield explicit formulae of $\Phi(q; S)$ and $\Psi(q; S)$. These, however, can be derived by converting the discrete-time, biased random-walk process into a continuous-time, drift diffusion model (DDM) (Ratcliff 1978). Indeed, Bogacz et al. (2006) have shown that as the time interval between successive samples goes to 0 (i.e., the sampling becomes continuous), the SPRT approaches the continuous-time DDM and it can be described by the following stochastic differential equation:

$$dV = Adt + cdW, V(0) = V_0 \quad (12)$$

where dV is the change in the decision variable over a small time interval dt . Equation 12 is comprised of 2 parts: 1) the constant drift Adt , which is the deterministic component of the process and describes the average increase in evidence supporting the correct

choice per time unit (in the OBC model, $A = I(q) = \frac{\log P(q|H)}{P(q|H^c)}$;

in the ABC model, $A = \delta(q - \theta)$ and 2) the white noise, cdW , which is the stochastic component of the process, a random, normally distributed, variable with mean 0 and variance c^2dt . In the OBC model, since $I(q)$ is a nonlinear function of q , c is itself a function of q and increases monotonically with respect to the squared deviation of q from the population mean (i.e., $c^2 \propto (q - \theta)^2 \sigma_{err}^2$). In contrast,

in the ABC model, since $u(q)$ is a linear function, c is independent of q : $c^2 = \delta^2 \sigma_{err}^2$.

Equation 12 can be converted to a Kolmogorov or Fokker-Plank equation and solved to find both the expected decision time and the expected error rate (Bogacz et al. 2006; Kimura and Moehlis 2012). More details on the approximation of SPRT to DDM can be found in Bogacz et al. (2006, pp. 745–746). In Appendix 2, I show the expression of Φ and Ψ as a function of drift, noise, and decision threshold. Derivations of these formulae using standard techniques of the first passage statistics of stochastic differential equations can be found in Gardiner (1985).

In the next section, I use Equations 9–12 to investigate how the expected benefits of a decision strategy S vary in relation to the costs of finding and assessing prospective mates. Since natural selection is expected to maximize the difference between the expected benefits and costs of choice, these analyses will provide insights into the functional role of parameterization biases.

RESULTS

In Figure 4, I compare the efficiency of OBC and ABC models, by assuming that ABC decision makers “know” with certainty both the parameters of the regression line. The comparison is carried out under 2 different “ecological” conditions. When quality assessment is accurate ($\sigma_{err}^2 = 1$ and $\sigma_{pop}^2 = 9$, Figure 4a,b), an efficient discrimination between lower-than-average and higher-than-average males is attained with a decision threshold set at $z = 4$. In this case, the ABC model performs slightly better than the OBC model in terms of discrimination ability (Figure 4a), but it takes much longer

to make decisions (Figure 4b). When quality assessment is poor ($\sigma_{err}^2 = 9$ and $\sigma_{pop}^2 = 9$), a similar discrimination power between lower-than-average and higher-than-average males is obtained by increasing the decision threshold (i.e., $z = 8$). Under these conditions, the OBC model performs better than the ABC model both in terms of discrimination power and decision time (Figure 4c,d).

The effect of searching and evaluation costs on optimal decision strategy

I use numerical methods to analyze how an optimal decision strategy varies with variation in the time-associated decision costs. I use the calibration coefficient k of Equation 11 to describe the weighting of costs relative to benefits. The expected difference between benefits and costs (i.e., the rewards) decreases monotonically with k , in both the OBC and ABC optimal decision strategies (Figure 5). Independent of k , however, the rewards provided by the OBC model are greater than those of the ABC model, although their differences slightly decrease as the costs increase. At very high costs (i.e., $k > 0.18$ in the ABC model and $k > 0.2$ in the OBC model), the OBC and ABC curves of optimal rewards converge to the line of no-choice strategy: under these conditions, evaluation costs are so high that the optimal decision strategy for a female is mating with the first male she encounters.

In Figure 6, I show how the optimal decision strategy changes with increasing costs. In the OBC model, both the quality threshold \bar{q}^* and the decision threshold z^* decrease with increasing costs, whereas V_0^* increases. The parameters of the optimal ABC strategy show a similar trend. The optimal value of the perceived population mean, \bar{q}^* , is positively biased under low evaluation costs and

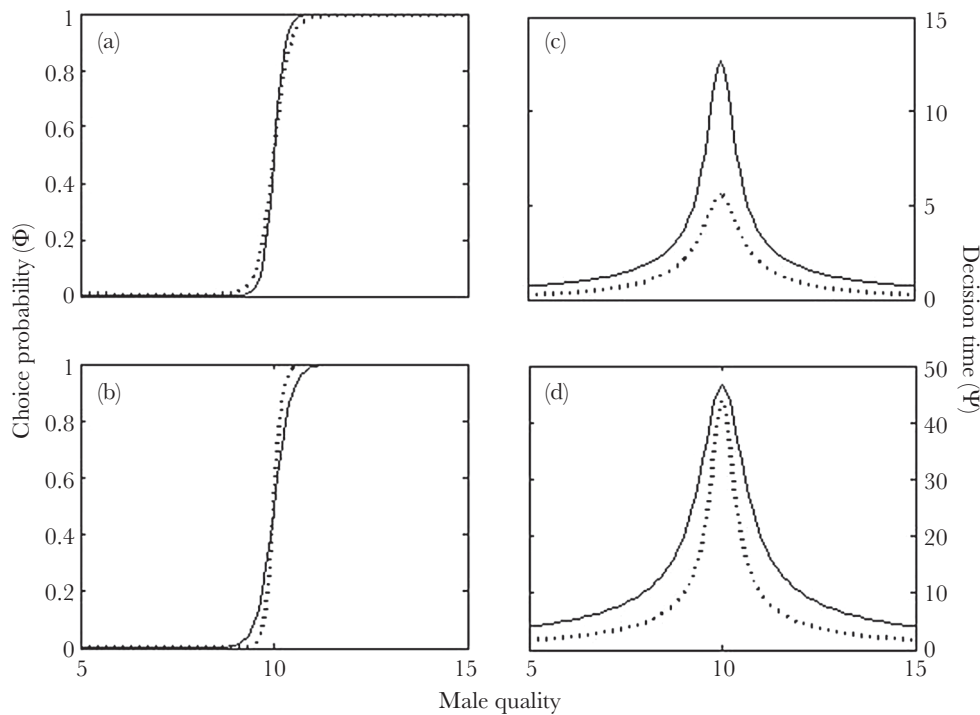


Figure 4

A comparison between the OBC (dotted line) and the ABC (solid line) model at 2 levels of assessment accuracy. The probability of choice and the mean time to make decision are plotted against the quality of the prospective mates. In (a) and (b), quality assessment is accurate ($\sigma_{err}^2 = 1$ and $\sigma_{pop}^2 = 9$) and the acceptance threshold is set at a low decisional threshold ($z = 4$), whereas in (c) and (d), it is highly inaccurate ($\sigma_{err}^2 = 9$ and $\sigma_{pop}^2 = 9$) and the acceptance threshold is set at $z = 8$. Under accurate evaluation, the ABC model performs slightly better than the optimal model in terms of choice probability, but it performs much worse in terms of decision time. Under inaccurate conditions, that is, in (c) and (d), the ABC model performs worse than the OBC model with respect to both the choice probability and decision time.

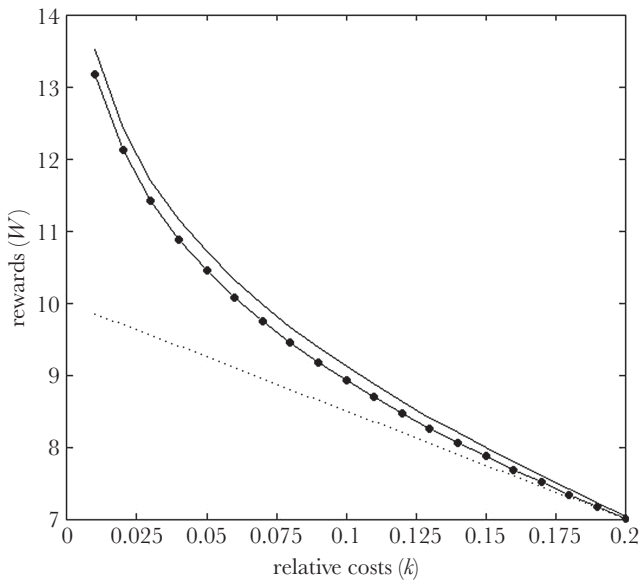


Figure 5

The effect of evaluation costs on the expected rewards of optimal decision strategy, in the OBC (solid line) and in the ABC model (full circles), compared with the expected rewards of the no-choice strategy (dotted line). The strength of the evaluation costs are described by the coefficient k , which is the weighting of costs relative to benefits (see Equation 11 in the main text). When the mate evaluation costs make mate choice advantageous, the OBC model always perform better than the ABC model. In these simulations, the parameters of the model are $\sigma_{\text{err}}^2 = 9$, $\sigma_{\text{pop}}^2 = 9$, $\theta = 10$.

negatively biased under high evaluation costs. This means that natural selection favors females that overestimate the expected quality of their prospective mates, when the searching costs are low, and that underestimate it when the costs are high. As predicted by the OBC model, in the optimal ABC decision strategies, q^* is positively correlated with z^* and negatively with V_0^* . Under low evaluation costs, the optimal decision threshold ($|z^*|$) is large, but the decision process starts closer to the negative boundary (i.e., $V_0^* < 0$), thus, making false-negatives (i.e., the rejection of males of quality $q > \theta$) much more likely than false-positives (i.e., the mating with a male of quality $q < \theta$). As the evaluation costs increase, the optimal decision threshold z^* decreases and V_0^* increases. At very high evaluation costs, V_0^* converges to the positive decision boundary, z^* . This represents the no-choice strategy: evaluation time is 0 and females mate with the first male they encounter, independent of his quality.

DISCUSSION

McNamara and Houston (2009) urged behavioral ecologists to a close integration between function and mechanism and suggested a research program (“evo-mecho”) specifically oriented to the systematic study of the evolution of mechanisms. The success of this research program, however, will largely depend on its capacity to develop a coherent body of theories and testable hypotheses. According to McNamara and Houston (2009), the research program should address questions that pertain to 3 levels of investigation. The first is concerned with the evolutionary explanations of the constraints imposed by rules (i.e., why rules are not completely flexible). The second level focuses on the optimal performance of rules and address the question of how natural selection has modified the parameters of a given rule to make it adapted to the specific requirements of the environment. The analysis of the functional role of cognitive biases pertains to this level and optimality or game-theory

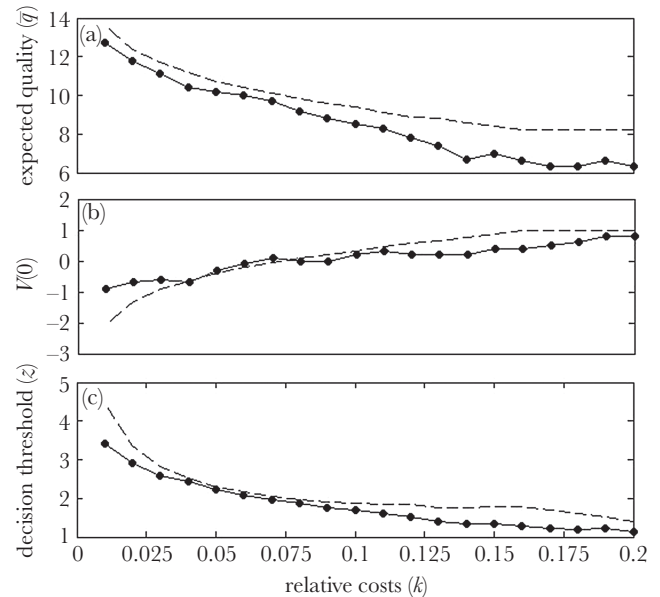


Figure 6

The effect of evaluation costs (k) on the optimal decision strategy in the OBC (dashed lines) and ABC (full-circle, solid lines) models. In these simulations, male quality is normally distributed with mean (θ) 10 and SD (σ_{pop}) 3, whereas the random error of mate-quality assessment (σ_{err}) is 3. The ABC strategies are defined by 3-parameters: 1) the mate expected quality (\bar{q}), which is positively biased ($\bar{q} > \theta$), when the assessment costs are low ($k < 0.6$) and negatively biased ($\bar{q} < \theta$), when the assessment are high (a); 2) the initial value of the decision variable (V_0), which is negatively biased (i.e., $V_0 < 0$) when the expected mate quality (θ) is overestimated, and positively biased (i.e., $V_0 > 0$) when it is underestimated (b); 3) the decision threshold (\bar{z}), which decreases as the evaluation costs increase (c). The effects of evaluation costs on the optimal ABC strategy are qualitatively similar to the effects they have on the optimal OBC strategy. In fact, the optimal ABC strategy shows a positive bias of \bar{q} and a negative bias of V_0 whenever the optimal OBC quality threshold, \bar{q} , is greater than θ . In these simulations, the parameters of the model are $\sigma_{\text{err}}^2 = 9$, $\sigma_{\text{pop}}^2 = 9$, $\theta = 10$.

models may represent the best theoretical tools to deal with these questions. Finally, the third level is concerned with the evolutionary significance of rules and of the organizational principles on which rules are based. In this paper, I have addressed decision making at both the second and the third level of analysis. To explore the evolutionary significance of decision rules (third level of analysis), I have proposed a 2-step optimality approach. The first step aims at identifying “computationally” optimal decision mechanisms. The second step aims at identifying “satisficing” approximations of these mechanisms. I have then used the optimal and the approximate-optimal model of decision making to address the second-level question, providing theoretical evidence for a functional role of cognitive bias.

Modeling the evolution of behavioral mechanisms: from theories of optimal computation to realistic hypotheses of suboptimal algorithms

The OBC model, which is the analogous of the SPRT (Wald 1945), satisfies the condition of computational optimality, because, for a given error rate (i.e., for a given accuracy level), it provides the fastest response (i.e., it minimizes decision time) (Wald and Wolfowitz 1948; Bogacz et al. 2006; McNamara et al. 2009). The “computationally” optimal model serves 2 important functions:

it identifies an upper limit on performance (McNamara and Houston 1980) and it helps deriving hypotheses on the actual decision mechanisms an animal uses, which is what the second step of the optimality approach is supposed to do. Indeed, from the OBC model, I have derived the ABC model, which approximates the optimal rule and which is supposed to model the actual decision mechanism.

Both the OBC and the ABC models assume that noisy evidence for the 2 alternative hypotheses is accumulated over time and that a decision is made when the resulting accumulated value reaches a critical threshold. In the last decade, neurobiological studies have given empirical support to these models (Bogacz 2007). For example, when a monkey is asked to choose between 2 visual stimuli, the firing rates of neuronal population associated to each alternative gradually increase up to a level at which decision is reached (Gold and Shadlen 2007). However, while in the OBC model, the increase in firing rate depends on the log-likelihood ratio of the 2 mutually exclusive hypotheses that the prospective mate is either an appropriate or inappropriate male, in the ABC model, it depends on the departure of the perceived quality from an expected standard. In fact, in the ABC model, a female is assumed to “know” only the average quality of prospective mates and the accuracy of her quality assessment. She obtains evidence for either one of the 2 hypotheses by representing their log-likelihood ratio in terms of a weighted deviation of the perceived male quality from the population mean. For this reason, the ABC algorithm is much less computationally demanding than the OBC algorithm and it could be easily implemented even in simple nervous systems. For example, the ABC model might explain the effect of social experience on mating preferences observed in the cricket *Teleogryllus oceanicus* (Bailey and Zuk 2008; 2009). In this species, females’ precopulatory experience influences mating behavior. In mate-choice experiments, females take longer to choose a mate when they have been previously exposed to male calling (Bailey and Zuk 2008; 2009) than when they have not. Moreover, females that have experienced low-attractive calls are more likely to choose a male and to retain his spermatophore than females that have experienced high-attractive calls (Rebar et al. 2011). The ABC model explains this plastic response to variation in social environment as the effect of differences in the parameters of the decision variable: females that have experienced highly attractive calls are predicted to be choosier than those exposed to low-attractive calls, because of their larger values of the expected mate quality (q).

As mentioned above, the proposed 2-step approach suggests a link between 2 questions that behavioral ecology has often considered as 2 distinct levels of analysis: the question about the mathematical procedure that can be used to find optimal solutions and the question about the mechanisms used to implement such solutions (McNamara and Houston 1980). In behavioral ecology, animals are often thought to follow simple “rules of thumb” tuned to the requirements of the environment by natural selection (Krebs and Davies 1993). A similar approach has been proposed in cognitive psychology with the “fast-and-frugal heuristics” research program (Goldstein and Gigerenzer 2002). Both approaches focus on rules, but neglect their underlying mechanisms and, by doing so, they implicitly assume that different rules are evolutionarily independent traits and that natural selection has provided decision makers with a “toolbox” of adaptive heuristics, each performing optimally (or at least satisfactorily) in different contexts.

Different rules, however, are not necessarily the expression of different mechanisms (Castellano 2009) and they may not be evolutionary independent traits. Natural selection acts on rules because it favors individuals that make the appropriate choice under the disparate conditions they experience in their life. Rules, however, are the expression of psychological and physiological mechanisms and natural selection is supposed to operate on them, sometimes favoring the evolution of new mechanisms, more often modifying mechanisms that already exist. Failing to understand the link between rules and mechanisms may represent a limit in our investigations of the functional significance of rules. To see why, consider, as an analogy, the evolution of bird wings. To understand how natural selection makes the bird forelimb adapted to flight, we do not necessarily need to know which was the forelimb anatomical structure of birds’ ancestors. However, if we do know the anatomy of the reptilian forelimbs, we could more easily understand what changes natural selection has promoted and why. The same argument holds for the evolution of behavioral rules. If we manage to recognize that different rules are the evolutionary variants of the same general mechanism, rather than evolutionary independent traits, we may more easily understand which parameters of the mechanism have been the target of selection and why. For this reason, in line with McNamara and Houston (2009) and Kacelnik (2012), I think that a coherent body of theory for the evo-mecho research program should be based on the characterization of these general mechanisms and the Marr’s approach is an effective way to pursue this goal (Castellano et al. 2012).

The ambiguous nature of adaptive cognitive biases

The ABC model suggests that there exist 2 kinds of cognitive biases in decision making: those that depend on the structure (representation bias) and those that depend on the parameters (parameterization bias) of the algorithm used to make decisions. The algorithm structure is likely to depend on the pattern of neural network connectivity and thus is probably relatively inflexible and under strong neurophysiological constraints. The algorithm parameters are more likely the effect of neuromodulation and they can be adjusted flexibly to the environmental conditions. We should expect behavioral flexibility to be due to changes in the algorithm parameters rather than to changes in algorithm structure.

As it concerns parameterization biases, the ABC model outlines some ambiguity in their characterization. For example, when a female overestimates the mean quality of prospective mates and, by doing so, she increases both the benefit (the mate expected quality) and the costs (time) of her choice, should we interpret the female’s behavior as the effect of an adaptive cognitive bias? From a mechanistic point of view, we should answer “yes.” The model assumes that the cognitive machinery is designed to estimate the probability that a mate has higher-than-average quality. If natural selection favors females that choose mates with quality, say, twice as high as average, then the cognitive machinery must be functionally adjusted and one way to do this is by introducing a systematic error in the decisional process. From a functional point of view, however, perhaps, our answer should be “no,” because if an increase in choosiness is strategically advantageous, then it is questionable whether to consider an error the way for attaining it. For example, the optimal decision model (OBC) predicts that for a decrease in the assessment cost, there should be an

increase in female choosiness, which can be attained by increasing both the reference standard (the male quality threshold) and the accuracy of the evaluation process (the decision threshold). In the ABC model, such a cost-dependent flexibility is attained by adjusting the parameters of the model in the direction predicted by the OBC model. Indeed, as the assessment costs decrease, not only females should improve their evaluation accuracy, but they should also overestimate the males' expected quality and make false-negative errors more likely than false-positive errors. Under the constraints imposed by the ABC model, the parameterization biases allow the decision process to perform as close as possible to the optimal solution.

The ambiguity in defining adaptive cognitive biases is not merely semantic, but it involves the role (and meaning) we assign to information in a communicative interaction (Rendall et al. 2009; Seyfarth et al. 2010). If we assume that decision makers do perceive the information conveyed by signals, then we may agree that a cognitive bias is what "skews our assessment away from an objective perception of information" (Johnson et al. 2013). However, receivers may respond to a signal not because they "know" what the signal means (what it refers to), but because they "know" how to react to it (Seyfarth and Cheney 2003). When a signal does convey information (because it associates to some undisclosed quality of signalers) but has no meaning (i.e., receivers do not use the signal to make an internal representation of the undisclosed quality of signalers), a cognitive bias in signal perception should be described, rather than as a departure from objectivity, as the mechanism that allows decision agents to adjust their decision rules to the internal and external conditions they are facing.

CONCLUSION

The integration between function and mechanism represents a fundamental step to unravel the functional role of behavioral flexibility. In this research program, cognitive biases play an important role because they may provide a window into the rules that govern behavioral flexibility. Behavioral rules, however, are the expression of psychological and physiological mechanisms and it is at this level of analysis that integration should occur. As argued by Kacelnik (2012, p. 37–38), "research into mechanisms of behavior should be closely associated to the constructs and methods of experimental psychology, rather than by the formulation of hypothetical rules of thumb or heuristics." Neurophysiological and psychological researches have provided support to the hypothesis that the process of decision making involves integration over time of noisy evidence and, thus, that it can be modeled as a statistical test (Bogacz 2007). Here, I have explored this hypothesis. I have modeled mate choice as a linear approximation of a sequential log-likelihood ratio test on the 2 alternative hypotheses that a prospective mate is either acceptable or unacceptable. But the same decision mechanism can be extended to all decision-making problems driven by perceptual events and it can be used not only in a sequential-choice context (one option at a time), but also in a simultaneous-choice context, when a decision maker is asked to choose simultaneously among several alternatives (Bogacz et al. 2006; McMillen and Holmes 2007). In this theoretical framework, the optimality approach still plays a central role, not only because it describes what is a rational choice, but also because it represents the starting point to formulate hypotheses on the mechanisms and rules that govern behavioral flexibility.

APPENDIX 1

The perceived utility

Let assume that the signalers' true quality (q) is normally distributed with a probability density function $f(q) = \mathcal{N}(\theta\sigma_{\text{pop}})$. The perceived quality of a signaler (i.e., his attractiveness Q) depends on his true quality, but in a stochastic rather than deterministic way. For simplicity, we define the perceived quality as:

$$Q = q + \varepsilon \quad (\text{A1})$$

ε is a normally distributed stochastic variable with 0 mean and standard deviation σ_{err} and is assumed to be independent of both signalers and receivers.

A male of quality q is perceived as having quality Q with probability:

$$P(Q | q) = \frac{f(q)}{\sqrt{2\pi}\sigma_{\text{err}}} e^{-\frac{(Q-q)^2}{2\sigma_{\text{err}}^2}} \quad (\text{A2})$$

Now, let assume that a male is acceptable if he has quality higher than the mean population male quality θ . The probability that an acceptable male is perceived as having quality Q is thus:

$$P(Q | q > \theta) = 1 - P(Q | q < \theta) = \int_{q=0}^{\infty} \frac{f(q)}{\sqrt{2\pi}\sigma_{\text{err}}} e^{-\frac{(Q-q)^2}{2\sigma_{\text{err}}^2}} dq$$

Since the evaluation error ε is unbiased (i.e., it has mean 0), it does not affect the expected perceived quality (Q) of acceptable and non-acceptable males, which can be obtained by computing the mean of the male-quality distribution truncated either above or below the population mean (Barr and Sherril 1999):

$$\begin{aligned} E[Q | q > \theta] &= \theta + 2z(0)\sigma_{\text{pop}} \\ E[Q | q < \theta] &= \theta - 2z(0)\sigma_{\text{pop}}. \end{aligned}$$

In contrast, the evaluation error ε does affect the variance of the quality of acceptable and nonacceptable males, which is:

$$\sigma_{Q|H}^2 = \sigma_{Q|H'}^2 = \sigma_{\text{err}}^2 + \sigma_{\text{pop}}^2(1 - 4z(0)^2)$$

Where z is the standardized normal probability density function (and $z(0) = 0.399$).

The probability density functions of acceptable and nonacceptable males differ from normal distributions being, respectively, positive or negatively skewed. However, when σ_{err}^2 and σ_{pop}^2 are of similar magnitude, we may approximate them to 2 normal distributions:

$$\begin{aligned} P(Q | q \geq \bar{q}) &= \frac{1}{\sqrt{2\pi(\sigma_{\text{err}}^2 + \sigma_{\text{pop}}^2(1 - 4z(0)^2))}} e^{-0.5 \frac{(Q - \bar{q} + 2z(0)\sigma_{\text{pop}})^2}{\sigma_{\text{err}}^2 + \sigma_{\text{pop}}^2(1 - 4z(0)^2)}} \\ P(Q | q < \bar{q}) &= \frac{1}{\sqrt{2\pi(\sigma_{\text{err}}^2 + \sigma_{\text{pop}}^2(1 - 4z(0)^2))}} e^{-0.5 \frac{(Q - \bar{q} - 2z(0)\sigma_{\text{pop}})^2}{\sigma_{\text{err}}^2 + \sigma_{\text{pop}}^2(1 - 4z(0)^2)}} \end{aligned}$$

The log-odds ratio $I(Q) = \frac{\log P(Q | q \geq \bar{q})}{P(Q | q < \bar{q})}$ may be approximated to the linear function:

$$u(Q) = -0.5 \frac{(Q - \bar{q} - 2z(0) \cdot \sigma)^2}{\sigma_e^2 + \sigma^2 (1 - 4z(0)^2)} + 0.5 \frac{(Q - \bar{q} + 2z(0) \cdot \sigma)^2}{\sigma_e^2 + \sigma^2 (1 - 4z(0)^2)} \tag{A3}$$

$$= \frac{4z(0)\sigma}{\sigma_e^2 + \sigma^2 (1 - 4z(0)^2)} (Q - \bar{q}).$$

By naming $d = 2z(0)\sigma$ and $\sigma_{QH}^2 = \sigma_e^2 + \sigma^2 (1 - 4z(0)^2)$, the Equation A3 becomes:

$$u(Q) = \frac{2d}{\sigma_{QH}^2} (Q - \bar{q}) \tag{A4}$$

APPENDIX 2

Consider the constant drift stochastic differential equation of Equation 12,

$$dV = Adt + cdW, \quad V(0) = V_0$$

with boundaries $\pm z$, $A = d(q - \bar{q})$ and $c = d\sigma_e$. By defining

$$\hat{q} = \frac{q - \bar{q}}{\sigma_e^2},$$

$$\hat{z} = \frac{z}{d},$$

and

$$\hat{V}_0 = \frac{V_0}{d}$$

it may be demonstrated (Gardiner 1985; cf. appendix of Bogacz et al. 2006, p. 747) that, for $q > \bar{q}$, the diffusion process is expected to cross the negative threshold ($-z$) (i.e., false-negative error, FN) with probability:

$$FN = \frac{1}{1 - e^{-\hat{z}\hat{q}}} - \frac{1 - e^{-2\hat{V}_0\hat{q}}}{e^{2\hat{z}\hat{q}} - e^{-2\hat{z}\hat{q}}} \tag{A5}$$

The probability that a female adopting a decisional strategy S_{ABC} will mate with a male of quality q is thus:

$$\phi(q; S_{ABC}) = 1 - FN \tag{A6}$$

Independent of the threshold, the average time of first passage statistics (i.e., the mean decision time) is:

$$\Psi(q; S_{ABC}) = \frac{\hat{z}}{(q - \bar{q})} \tanh(\hat{z}\hat{q}) + \frac{2\hat{z}}{e^{2\hat{z}\hat{q}} - e^{-2\hat{z}\hat{q}}} (1 - e^{-2\hat{V}_0\hat{q}}) \tag{A7}$$

Equations A5 and A7 show that among the 4 parameters controlling the decision process (i.e., z , V_0 , q , and d), one (i.e., the perceived accuracy, d) is redundant. In fact, if a decision maker is forced to change this parameter of a quantity b (i.e., $d' = d + b$), she can compensate the effect of this change by correcting the decision threshold, z , and the initial condition, V_0 , by a factor $(d + b)/d$, so that \hat{z} is maintained constant.

FUNDING

University of Turin (ex 60%) (CASSRILO13).

I thank 3 anonymous reviewers for their constructive criticisms.

Handling editor: Shinichi Nakagawa

REFERENCES

Bailey NW, Zuk M. 2008. Acoustic experience shapes female mate choice in field crickets. *P Roy Soc B-Biol Sci.* 27:2645–2650.

Bailey NW, Zuk M. 2009. Field crickets change mating preferences using remembered social information. *Biol Lett.* 5:449–451.

Barr DR, Sherril TE. 1999. Mean and variance of truncated normal distributions. *Am Statist.* 53:357–361.

Bateson M, Desire S, Gartside SE, Wright GA. 2011. Agitated honeybees exhibit pessimistic cognitive biases. *Curr Biol.* 21:1070–1073.

Bateson M, Healy SD. 2005. Comparative evaluation and its implications for mate choice. *Trends Ecol Evol.* 20:659–664.

Bateson M, Healy SD, Hurly TA. 2003. Context-dependent foraging decisions in rufous hummingbirds. *P Roy Soc B-Biol Sci.* 270:1271–1276.

Bentosela M, Jakovcevic A, Elgier AM, Mustaca AE, Papini MR. 2009. Incentive contrast in domestic dogs (*Canis familiaris*). *J Comp Psychol.* 123:125–130.

Bogacz R. 2007. Optimal decision-making theories: linking neurobiology with behaviour. *Trends Cog Sci.* 11:118–125.

Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD. 2006. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psych Rev.* 133:700–765.

Bowers JS, Davis CJ. 2012. Bayesian just-so stories in psychology and neuroscience. *Psychol Bull.* 138:389–414.

Castellano S. 2009. Towards an information-processing theory of mate choice. *Anim Behav.* 78:1493–1497.

Castellano S, Cadeddu G, Cermelli P. 2012. Computational mate choice: theory and empirical evidence. *Behav Process.* 90:261–277.

Fawcett TW, Fallenstein B, Higginson AD, Houston AI, Mallpress DEW, Trimmer PC, McNamara JM. 2014. The evolution of decision rules in complex environments. *Trends Cogn Sci.* 18:153–161.

Fawcett TW, Hamblin S, Giraldeau LA. 2013. Exposing the behavioral gambit: the evolution of learning and decision rules. *Behav Ecol.* 24:2–11.

Gardiner CW. 1985. *Handbook of stochastic methods.* 2nd ed. New York: Springer.

Gigerenzer G, Czerlinski J, Martignon L. 2002. How good are fast and frugal heuristics. In: Gilovitch T, Griffin D, Kahneman D, editors. *Heuristics and biases, the psychology of intuitive judgment.* Cambridge: Cambridge University Press. p. 559–581.

Gilovitch T, Griffin D, Kahneman D. 2002. *Heuristics and biases, the psychology of intuitive judgment.* Cambridge: Cambridge University Press.

Gold JI, Shadlen MN. 2007. The neural basis of decision making. *Annu Rev Neurosci.* 30:535–574.

Goldstein DG, Gigerenzer G. 2002. Models of ecological rationality: the recognition heuristic. *Psychol Rev.* 109:75–90.

Griffiths TL, Chater N, Norris D, Pouget A. 2012. How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012). *Psychol Bull.* 138:415–422.

Johnson DDP, Blumstein DT, Fowler JH, Haselton MG. 2013. The evolution of error: error management, cognitive constraints, and adaptive decision-making biases. *Trends Ecol Evol.* 28:474–481.

Johnson DDP, Fowler JH. 2011. The evolution of overconfidence. *Nature.* 477:317–320.

Kacelnik A. 2012. Putting mechanisms into behavioral ecology. In: Hammerstein P, Stevens JR editors. *Evolution and the mechanisms of decision making.* Cambridge (MA): MIT Press.

Kimura M, Moehlis J. 2012. Group decision-making models for sequential tasks. *SIAM Rev.* 54:121–138.

Krebs JR, Davies NB. 1993. *An introduction to behavioral ecology.* 3rd ed. Oxford: Blackwell.

Lange A, Dukas R. 2009. Bayesian approximations and extensions: optimal decisions for small brains and possibly big ones too. *J Theor Biol.* 259:503–516.

Latty T, Beekman M. 2011. Irrational decision-making in an amoeboid organism: transitivity and context-dependent preferences. *P Roy Soc B-Biol Sci.* 278:307–312.

Marr D. 1982. *Vision: a computational investigation into the human representation and processing of visual information.* New York: Freeman.

- Marshall JAR, Trimmer PC, Houston AI, McNamara JM. 2013. On evolutionary explanations of cognitive biases. *Trends Ecol Evol.* 28:469–473.
- McMillen T, Holmes P. 2007. The dynamics of choice among multiple alternatives. *J Math Psychol.* 50:30–57.
- McNamara JM, Fromhage L, Barta Z, Houston AI. 2009. The optimal coyness game. *P Roy Soc B-Biol Sci.* 276:953–960.
- McNamara JM, Houston AI. 1980. The application of statistical decision theory to animal behaviour. *J Theor Biol.* 85:673–690.
- McNamara JM, Houston AI. 2009. Integrating function and mechanism. *Trends Ecol Evol.* 24:670–675.
- Parker GA, Smith JM. 1990. Optimality theory in evolutionary biology. *Nature.* 348:27–33.
- Ratcliff R. 1978. A theory of memory retrieval. *Psychol Rev.* 85:59–108.
- Rebar D, Zuk M, Bailey NW. 2011. Mating experience in field crickets modifies pre and postcopulatory female choice in parallel. *Behav Ecol.* 22:303–309.
- Rendall D, Owren MJ, Ryan MJ. 2009. What do animal signals mean? *Anim Behav.* 78:233–240.
- Seyfarth RM, Cheney DL. 2003. Signalers and receivers in animal communication. *Ann Rev Psychol.* 54:145–173.
- Seyfarth RM, Cheney DL, Bergman T, Fischer J, Zuberbuhler K, Hammerschmidt K. 2010. The central importance of information in studies of animal communication. *Anim Behav.* 80:3–8.
- Shafir S, Waite TA, Smith BH. 2002. Context-dependent violations of rational choice in honeybees (*Apis mellifera*) and gray jays (*Perisoreus canadensis*). *Behav Ecol Sociobiol.* 51:180–187.
- Tversky A, Kahneman D. 1974. Judgment under uncertainty: heuristics and biases. *Science.* 185:1124–1131.
- Waite TA. 2008. Preference for oddity: uniqueness heuristic or hierarchical choice process? *Anim Cogn.* 11:707–713.
- Wald A. 1945. Sequential tests of statistical hypotheses. *Ann Math Statist.* 16:117–186.
- Wald A, Wolfowitz J. 1948. Optimum character of the sequential probability ratio test. *Ann Math Statist.* 19:326–339.