

Artificial life, natural rationality and probability matching

Benoit Hardy-Vallée

Department of Philosophy, University of Waterloo
200 University Avenue West, Waterloo, Ontario
Canada N2L 3G1

Abstract-Three different agendas for the philosophy of ALife are presented. While the metaphysical and the epistemological ones have been extensively developed, a third one is proposed here. The naturalistic agenda is more interested in resolving empirical problems. Consequently, this paper addresses an empirical issue: the adaptive function of a phenomenon known as “probability matching”, that is, the tendency to match the probability of choice with the probability of reward. Probability matching and its common interpretations are presented and discussed. Based on ALife and Neural Networks simulations, I present an alternative hypothesis: probability matching is an adaptation to radical uncertainty.

I. INTRODUCTION

Most discussions of Artificial Life (ALife hereafter) fall under two categories. In the first one, the philosophy of ALife focuses on the reliability of ALife modeling as a scientific tool, and reflects on whether simulations and robotic experiments are deductive/inductive instruments, theoretical biology research, scientific models, thoughts experiments, *etc.* This could be called the *epistemological agenda*. In the second one, philosophical inquiry also deals with the role of ALife in the definition and understanding of important notions: life, cognition, emergence, evolution or adaptation. This could be construed as the *metaphysical agenda*. These two agendas [1-3], however important, push the philosophy of ALife toward the study of broad and deep questions such as the nature of scientific explanations or the relation between life and cognition. This attitude implies that philosophy and ALife are two separate fields and that philosophy uses ALife to validate or invalidate its theories. This conception of interdisciplinary relationship fits the pre-Quinean picture of philosophy: while science deal with the empirical world, philosophy is a meta-theoretical discipline that analyses abstracts objects like propositions, theories, or norms. Nevertheless, after Quine, philosophers begins to conceive their discipline as continuous with science: concepts from both fields are so interwoven that there is no strict boundaries between them [4]. Naturalism, according to Quine, implies that philosophy and science are partner on the same epistemic project: the description of reality. Since there is no principled reason why philosophical activity should be restrained to epistemological or metaphysical questions and the practice of naturalistic philosophy is already well established, I suggest that the philosophy of ALife should set upon another project that I would call the *naturalistic agenda*.

In this conception, the philosophy of ALife is not only about broad questions or conceptual definitions, but is on par with ALife and other philosophical and scientific domains (philosophy of mind, cognitive science, neuroscience, *etc.*) in the “pursuit of truth”[5]. I therefore advocate a Quinean approach to ALife, in which ALife *cum* other fields helps us in resolving problems that may be relevant to ALife or not. (Note that I do not pretend that Quinean approaches to ALife do not already exist, but only that they are more peripheral in the philosophy of ALife).

One project in which the naturalistic agenda may bear fruit is in what I call “natural rationality” [6]. Natural rationality is the study, from both a descriptive and a normative point of view, of the cognitive mechanism by which humans or other animals make decisions. Most approaches of rationality take only the descriptive or the normative side, and hence tend either to describe cognitive/neuronal processes without concern with their optimality [7], or to state *a priori* conditions for rational behavior [8]. Natural rationality is an attempt to bridge the gap between these two projects without the *a priori* biases commonly found in either side of the debate on normative/descriptive issues (classical rationality is not natural, rationality cannot be described in scientific terms, *etc.*). It is not an epistemological or metaphysical project, but a more limited one concerned with the structure and evolution of decision-making processes found in nature. In this paper, I am interested in the phenomenon known as “probability matching” (PM hereafter), the tendency to match the probability of choice with the probability of reward. In section II, I introduce PM and its common interpretations. Section III present an ALife modeling that suggests an evolutionary scenario for PM while section IV argue that PM should be understood as a rational strategy to cope with *radical* uncertainty.

II. PROBABILITY MATCHING

Probability matching (PM) is a widely observed phenomenon in which subjects match the probability of choices with the probability of reward in a stochastic context. For instance, suppose one has to choose between two sources of reward: one (A) that gives reward on 70% of the occasions, and the other (B) on 30%. The rational, utility-maximizing strategy is to choose always A. The matching strategy consists in choosing A on 70% of the occasions and B on 30% of the occasions. While the former leads to a reward 7 times out of 10, the latter will be rewarding only 5.8 times out of 10 [(0.7 x

0.7) + (0.3 x 0.3) = 0,58]. Clearly, the maximizing strategy outperforms the matching strategy.

The maximizing strategy, however, is rarely found in the biological world. From bees to birds to humans, most animals match probabilities [9-13]. In typical experiments with humans, subjects are asked to predict which light will flash (left or right for instance) and have a monetary reward for every correct answer. Rats have to forage for food in a T-maze, pigeons press levers that reward food pellets of different size with different probability, while bees forage artificial flowers with different sucrose delivery rate. In all cases, the problem amount to efficiently maximize reward from various sources, and the most common solution is PM. (There are variations, but PM predicts reliably subjects' behavior). Different probability distributions, rewards or context variations do not significantly alter the results. Hence it is a particularly robust phenomenon, and a clear example of a discrepancy between standards of rationality and natural behavior. Three different perspectives could then be adopted: 1) subjects are irrational, 2) subjects are boundedly rational and hence cannot avoid such mistakes or 3) subjects are in fact ecologically rational and hence PM is not irrational.

According to the first one, mostly held in traditional normative economics and decision theory [e.g., 14], this behavior is blatantly irrational. Rational agents rank possible actions according to the product of the probability and utility of the consequences of actions, and they choose those that maximize subjective expected utility. In opting for the matching strategy, subjects violate the axioms of decision theory, and hence their behavior cannot be rationalized. In other words, their preferences cannot be construed as maximizing a utility function: it is "an experimental situation which is essentially of an economic nature in the sense of seeking to achieve a maximum of expected reward, and yet the individual does not in fact, at any point, even in a limit, reach the optimal behavior" [15].

Another perspective, found in the "heuristic and biases" tradition [16, 17] also considers that it is irrational but suggests why this particular pattern is so common. The boundedly rational mind cannot always proceed to compute subjective expected utilities but rely on simplifying tricks: heuristics. One heuristic that may explain human shortcomings in this case is representativeness: judging the likelihood of an outcome by the degree to which it is representative of a series. This is how the phenomena known as the gambler's fallacy (the belief that an event is more likely to occur because it has not happened for a period of time) may be explained: "there was five heads in a row; there *cannot* be another one!" This heuristics may also explain why subjects match probabilities: it is more likely that if the 70% source was rewarding in the last round, it would be better to try the 30% a little in order to maximize reward. Hence PM is irrational, but this irrationality is excusable, albeit without any particular significance.

The third perspective, that could be either named "ecological rationality" or "evolutionary psychology" [18-21] argue

instead that humans and animals are not really irrational, but adapted to certain ecological conditions whose absence explains apparent irrationality. Ecologically rational heuristics are not erroneous processes, but mechanisms tailored to fit both the structure of the environment and the mind: they are fast, frugal and smart. PM can be rational in some context and irrational in another one. For instance, when animals are foraging and competing with conspecifics for resources, PM is the optimal strategy:

(...) if one considers a natural environment in which animals are not as socially isolated as in a T-maze and in which they compete with one another for food, the situation looks different. Assume that there are a large number of rats and two patches, left and right, with an 80:20 distribution of food. If all animals maximized on an individual level, then they all would end up in the left part, and the few deviating from this rule might have an advantage. Under appropriate conditions, one can show that probability matching is the more rational strategy in a socially competitive environment [22].

This pattern of behavior and spatial distribution correspond to the Ideal Free Distribution (IFD) model used in behavioral ecology [23]. Derived from optimal foraging theory [24], the IFD predicts that the distribution of individuals between food patches will match the distribution of resources, a pattern observed in many occasions in animals and humans [25-29].

There are of course discrepancies between the model and observed behaviors, but foraging groups tend to approximate the IFD. This supports the claim that PM is a rational heuristic only in a socially competitive environment: it could also be construed as a mixed-strategy Nash equilibrium in a multiplayer repeated game [30] or as an evolutionarily stable strategy, that is, a strategy that could not be invaded by another competing strategy in a population who adopt it [31]. Seth's simulations [32] showed that a simple behavioral rule may account for both individual and collective matching behavior.

Are we justified here in concluding that, at the *individual* level, PM is suboptimal and irrational? Although the common wisdom in evolutionary psychology and ecological rationality is affirmative [7, 33, 34], I will argue that, while PM is an efficient social heuristics, it is also an efficient, adaptive, *individual* heuristics in some contexts. This account will be supported by findings in artificial life, neural networks simulations and neuroscience.

III. THE (ARTIFICIAL) EVOLUTION OF PROBABILITY MATCHING

In a recent study, Niv, Joel, Meilijson and Ruppin [35; NJMR hereafter] showed that PM evolves when the environment is highly uncertain. NJMR's model simulated the evolution of bees in a foraging environment. Instead of handcrafting the synaptic learning rule or evolving only synaptic weights or architectures, NJMR attempted to evolve the network's learning rule.

The artificial bees forage a (virtual) 3D space where a portion of the ground is a 60 x 60 grid of blue and yellow

squares (see Fig. 1a). Each square represents a flower, and each color is a kind of flower that has a particular reward contingency, randomly modified at every generation *and* during the bee's life. A particular flower type (blue or yellow) may indicate either a constant $0.7 \mu\text{l}$ nectar reward, or a $1.0 \mu\text{l}$ nectar in 20% of the occasions and 0 otherwise. The bee's "life" begins on a randomly assigned location, consists in moving from one flower to another while maximizing nectar intake, and ends after 100 of such trials. Its vision field is a 10° cone (see Fig. 1a) and the bee can, at each step, either maintain its orientation or reorient it randomly. The bee can go to any adjacent flower from where it is currently located after consuming all available nectar in the flower. This task is not an easy one: successfully foraging an environment is a problem similar to the Traveling Salesperson [36], a problem known to be NP-complete. Thus foraging strategies must be efficient while simple, because the optimal solution involves complex computational capacities.

The bee's brain is composed of three modules: 1) the "regular" specify the percentage of the yellow, blue and neutral patches in the visual field, 2) the "differential" indicate the temporal difference of these values (e.g., the difference between the actual percentage of yellow and the previous one), and 3) the "reward" module report the quantity of nectar found in the flower (see Fig. 1b). All modules are connected to a linear neuron P [mirroring the functioning of the VUMmx1 neuron in real bees, cf. 37], whose output determines the bee's action.

The bee's genome is composed of 28 genes coding for a parameter of the neural network architecture or dynamics: for instance, 15 code for the "innate" structure of the network; 7 determine the existence or inexistence of each synapse; 6 code for the initial weight of the regular and differential module while the others specify the parameters of action and learning dynamics. A first generation of 100 bees was produced by generating random genomes. These bees foraged the environment and ended their life with a fitness score proportional to their average nectar intake per trial. That fitness score also determined their probability of reproducing. A genetic algorithm optimized the bees' performance by forming 50 pairs of parents (matched according to their fitness score), combining their genome after mutation and crossover and producing 2 offspring per pair of parents. The offspring went foraging and the evolutionary process was repeated for 500 generations.

Although not all evolutionary runs were successful, those who were all ended up with a network architecture similar to Montague *et al.* model [38]: a dependency of the differential module on the reward module and a modification of these intermodule synapses correlated to the presence of reward. Selected individuals exhibited efficient reinforcement learning and developed a preference for the constant $0.7 \mu\text{l}$ nectar reward, replicating observation of real bees in similar context [39]. When reward contingencies changed (switching between blue and yellow), the network's dynamics adjusted itself to the

new payoff structure. Moreover, when implemented in a Kephra mobile robot, the neural network architecture led to similar behavior, suggesting that the embodied nature of foraging and learning was well represented in the simulation. Hence this simulation led to realistic learning rules and behaviors.

An important finding of this simulation is that bees naturally developed a strategy close to PM. While other simulations [11], in line with the conventional justification of PM, construe this phenomenon as an outcome of competitive foraging, NJMR's model indicate that matching probability may emerge out of the evolutionary and reinforcement dynamics of agents situated in complex environment. This makes sense if one considers that PM is an adaptive strategy for radically uncertain environment.

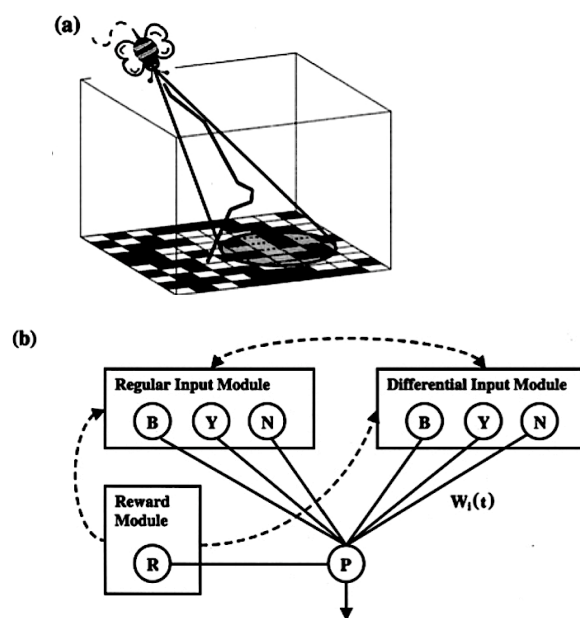


Fig. 1, from Niv et al (2002). a) schematic representation of the bee and its environment. b) the neural network of the bee (B=blue flowers, Y=yellows, N=neutral, R=reward)

IV. AN ADAPTIVE HEURISTICS FOR RADICAL UNCERTAINTY

Most of discussions of PM assume that probabilities are known when the decision is taken. Hence when it is asserted that animals or humans match probabilities, it is often implied that they do that in a context of *first-order uncertainty*: knowing that A will be rewarding in 70% of the occasions is uncertain knowledge because one do not know for sure what will be the next outcome (one can only know that there is a 70% probability that it is a reward). In some situations however, uncertainty can be *radical*, or *second-order uncertainty*: even the probabilities are unknown. When the assumption of known probabilities is dropped, the irrationality of PM is less obvious.

Under radical uncertainty, cognitive agents must learn reward probabilities. Learners in this setting must, at the same time, explore their environment in order to gather information about its payoff structure and exploit this information to obtain reward. They face a deep problem – known as the exploration/exploitation trade-off – because they cannot do both at the same time: one cannot explore all the time, one cannot exploit all the time and exploration must be reduced but cannot be eliminated. A well-known example of this trade-off is *K-armed bandit problem*.

Suppose an agent has n coins to spend in a slot machine with K arms (here $K=2$ and we will suppose that one arm is high-paying and the other low-paying, although the agent does not know that). The only way the agent has access to the arms' rate of payment – and obtains reward – is by pulling them. Hence she must find an optimal trade-off when spending its coins: trying another arm just to see how it pays or staying with the one who already paid? The goal is not only to maximize reward, but also to maximize reward *while obtaining information about the arm's rate*. Two types of error may appear: false negative (a low-paying sequence of the high-paying arm) or false positive (a high-paying sequence of the low-paying paying arm).

To solve this problem, the optimal solution is to compute an index for every arm, updating this index according to the arm's payoff and choosing the arm that has the greater index [40]. In the long run, this strategy amounts to following decision theory after a learning phase. But as soon as switching from one arm to another has a cost, as Banks & Sundaram [41] showed, the index strategies cannot converge towards an optimal solution. A huge literature in optimization theory, economics, management and machine learning addresses this problem [42-44]. These researches look for the normative, optimal policy to adopt in K -armed bandit problems. Studies of humans or animals explicitly submitted to bandit problems, however, show that subjects tend to rely on the matching strategy [45]. In one study, for instance, [46], subjects were required to select between two icons displayed on a computer screen; after each selection, a slider bar indicated the actual amount of reward obtained. The matching strategy predicted the subject's behavior, and the same results hold for monkeys in a similar task [47, 48].

I suggest that the pervasiveness of PM in nature is neither an accident nor a spandrel [49], and propose the following hypothesis: PM is an adaptation to uncertainty. Put differently, PM is a trait that was selected for its fitness-enhancing quality in uncertain environment. It is not the *perfect* strategy for coping with uncertainty but given the developmental, computational and ecological constraints that living beings face, this strategy is a “good enough” one that beats others. The maximizing strategy (choosing only the highest probable source of reward) is not the best one because it does not provide a robust reward intake across contexts.

As the last section showed, NJMR's simulated bees where situated in a bandit-like environment: they did not know the

probabilities of reward and these probabilities changed at some point. Thus, in this environment, when cognitive systems able to record information about reward can differentially reproduce, natural selection leads to generations of individuals that match probabilities.

A reasonable hypothesis about the neurological underpinning of this trait is that mechanisms guiding matching behavior are reinforcement learning algorithms realized by the brain's reward center. NJMR's evolved neural architecture, replicating Montague *et al.*'s architecture [38], implements a learning rule analogous to TD-learning, a unsupervised learning rule that update reward estimation according to prediction error. Systems implementing TD-learning use the difference between sequential predictions to learn to predict accurately: whenever an unpredicted reward or absence of predicted reward is detected, TD algorithms revise prediction. Artificial Neural Networks and robots implementing TD-learning reproduced humans, primate and bees behavior [48, 50-52]. Neuroscience indicate that the TD mechanism are realized in neuromodulatory processes such as dopaminergic systems in vertebrates and octopaminergic systems in invertebrates [53]. Dopaminergic/octopaminergic systems provide homogeneous response and act as a common currency for a reward signal [54, 55]. One of the interesting feature of these neurons is that they also code for probability matching [48]. Therefore it is plausible that PM emerges of the learning dynamics of natural brains. Nevertheless, to be considered as an adaptation, the advantages of PM should be demonstrated.

One must first note that PM is a simple heuristic: instead of tedious computations of indices, PM records the performance of an action with its rate of success and recall it when needed. Even simple systems like bees may rely on it. The only cognitive representation involved in action selection would be the reward rate: the system needs only to map this observed rate to motor planning. It is not the best policy available, but it is cognitively undemanding. Moreover, PM can be optimal in at least 3 circumstances.

First, when probabilities are identical across reward sources (ex: 50-50, or 20-20-20-20), first-order uncertainty is maximal and in this case only is equivalent to second-order uncertainty: knowing that it will either rain or not tomorrow is tantamount to not knowing what will be the weather. In this case, PM, decision theory and Ideal Free Distribution all agree: the best solution is to sample every source in an equal proportion (unless it is easier to stick with one resource). Thus PM is optimal in this case.

A second scenario where PM is optimal is when probabilities are unknown: in second-order uncertainty, every option is *a priori* equivalent. For instance, in a 2-armed bandit problem without any prior knowledge, the reward probability of A and B must be initially set to 50% and hence the optimal strategy is to try each option half of the time. Hence PM is also optimal in this case.

A third case where PM is advantageous is in readjusting probability in a changing environment. For instance, in the

CONCLUSION

NJMR's model, foraging bees may learn that one flower type (say blue) has a 0.7 μl nectar reward on all occasions, and the other type has 1 μl nectar in 20% of the occasions. The optimal policy, according to decision theory, would be to collect nectar only from the former. However, if the reward contingencies of the flowers change, sticking to a strict policy such as "collect nectar only on blue flowers" may be suboptimal if blue flowers become the less rewarding flower. If the agent follows a PM-style policy, the change in the flower reward rate will induce a change in the visit rate. In this setting, the PM strategy outperforms a rigid policy. NJMR's artificial bees demonstrated the adaptivity of PM, and switched their preferences when reward contingencies changed. This example highlights the optimality of PM in dynamic context. As Goldberg noted, "if the environment of decision is changing such that it is possible for the environment to go against our current thinking, then probability matching can hold the upper hand during the readjustment phase." [56]. The problem of most of the solutions to armed bandit is that they assume the stationarity of the environment. If, however, the environment is non-stationary, PM is a better strategy. Here again, PM is optimal.

Therefore, PM is a policy adequate for radical uncertainty: whether this uncertainty comes from probabilistic distribution, lack of information or environmental change, it is a rational policy to adopt and it is found in vertebrate and invertebrate nervous systems. Besides being efficient in uncertain contexts, it also efficient in social foraging (see section II), hence this suggests that PM is an adaptive heuristic that may be useful in a lot of settings, individual or not. It does not imply that PM is always rational: when probabilities are known and the environment is stationary, the best strategy is still the maximizing one. Yet the difficulty with this situation is that it is an idealization and thus rarely occurs. Therefore, maximizing is a solution for a problem almost nothing on Earth faces, because one may never know whether the observed probabilities are adequate and refer to stationary processes. Signal detection [57], natural resources, weather, climate, competition and the complex economy of biological markets [58, 59] are sources of uncertainty. There are of course exceptions: in the course of human history, we invented statistics and probability theory, and we must reason sometimes as if our probabilities represent accurately fixed properties. In these formal contexts, maximizing is the rational policy and behavioral studies indicate that it may emerge when certain conditions are present: greater cognitive abilities of the subjects, [60] larger incentives, feedback and training [61]. Thus one must distinguish contexts where PM is really suboptimal from contexts where it is a good heuristics for coping with uncertainty. It is irrational in context where information is – or should be – considered to be complete and reliable.

The philosophy of ALife does not need to focus only on epistemological and metaphysical issues. Questions more limited in scope may figure in a *naturalistic* agenda, where ALife and other disciplines are woven together. Natural rationality, the research and evaluation of decision-making mechanisms, is an instance of naturalistic inquiry that integrate ALife. Hence the present paper focused on the nature and adaptive function of probability matching. Contrarily to claims according to which PM is either irrational, a by-product of heuristic and biases or only rational in competitive foraging, I argued that PM must be construed as an adaptation to uncertainty. Neuromodulatory processes in nervous systems induce TD-learning, a robust and efficient learning rule that tends to match probabilities, and PM's efficacy was demonstrated in second-order uncertainty contexts. NJMR simulated the evolution of foraging bees in bandit-like environments and showed how evolution and development lead to PM. This, plus the fact that PM is both ubiquitous in nature and efficient in uncertain contexts, supports the hypothesis of PM as an adaptive heuristic that evolution and learning bring about. Thus in this sense, PM could be considered as an evo-devo adaptation. Because life is often a big K-armed bandit problem, evolution selects developmental systems that adjust themselves efficiently to their radically uncertain environment:

A system that learns through prediction learning need not have the path from goal to reward specified, in contrast to fixed behavioral patterns, such as stimulus-response learning. Instead, the path from goals to rewards may be left open and discoverable via learning, resulting in flexible action. Evolution, then, may shape the pattern of basic rewards animals are motivated to obtain, but the behavioral path is left open to discovery, as are more complex relations among predictors. [62]

Although the behavioral path is left open to discovery, nervous systems have the propensity to discover PM generations after generation because of its simplicity, naturalness and usefulness. This hypothesis is more general than the strict Ideal Free Distribution explanation of PM, according to which its only adaptive advantage is in competitive foraging, and hence treats competitive foraging as a specific kind of uncertainty problems solved by PM. Further simulations and empirical work will be needed to test this hypothesis.

REFERENCES

- [1] D. C. Dennett, "Artificial Life as Philosophy," *Artificial Life*, vol. 1, pp. 291-292, 1994.
- [2] M. A. Boden, *The philosophy of artificial life*. Oxford ; New York: Oxford University Press, 1996.
- [3] Bedau, "Philosophical Content and Method of Artificial Life," in *The Digital Phoenix: How Computers are Changing Philosophy*, T. W. Bynam and J. H. Moor, Eds.: Blackwell, 1998, pp. 135-152.
- [4] W. V. O. Quine, "Two Dogmas of Empiricism," *The Philosophical Review*, vol. 60, pp. 20-43, 1951.
- [5] W. V. O. Quine, *Pursuit of truth*. Cambridge, Mass.: Harvard University Press, 1990.
- [6] B. Hardy-Vallée, "Brains and decisions in the economy of nature," in *Companion to Contemporary Philosophy of Biology*, G. Terzis and R. Arp, Eds. Aldershot (U.K.): Ashgate, forthcoming, 2007.
- [7] G. Gigerenzer, "Fast and frugal heuristics: The tools of bounded rationality," in *Blackwell handbook of judgment and decision making*, D. Koehler and N. Harvey, Eds. Oxford: Blackwell, 2004, pp. 62-88.
- [8] D. Davidson, *Problems of rationality*. Oxford: Oxford University Press, 2004.
- [9] C. R. Gallistel, *The organization of learning*. Cambridge: MIT Press, 1990.
- [10] I. Erev and G. Barron, "On adaptation, maximization, and reinforcement learning among cognitive strategies," *Psychol Rev*, vol. 112, pp. 912-31, 2005.
- [11] A. K. Seth, "Modeling Group Foraging: Individual Suboptimality, Interference, and a Kind of Matching," *Adaptive Behavior*, vol. 9, pp. 67-89, 2001.
- [12] N. Vulkan, "An Economist's Perspective on Probability Matching," *Journal of Economic Surveys*, vol. 14, pp. 101-118, 2000.
- [13] U. Greggers and R. Menzel, "Memory dynamics and foraging strategies of honeybees," *Behavioral Ecology and Sociobiology*, vol. V32, pp. 17-29, 1993.
- [14] L. J. Savage, *The foundations of statistics*. New York.: Wiley, 1954.
- [15] K. J. Arrow, "Utilities, Attitudes, Choices: A Review Note," *Econometrica*, vol. 26, pp. 1-23, 1958.
- [16] D. Kahneman, P. Slovic, and A. Tversky, "Judgment under uncertainty : heuristics and biases." Cambridge ; New York: Cambridge University Press, 1982, pp. xiii, 555.
- [17] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, pp. 263-291, 1979.
- [18] J. H. Barkow, L. Cosmides, and J. Tooby, *The Adapted mind : evolutionary psychology and the generation of culture*. New York: Oxford University Press, 1992.
- [19] G. Gigerenzer, *Adaptive thinking : rationality in the real world*. New York: Oxford University Press, 2000.
- [20] G. Gigerenzer, P. M. Todd, and ABC Research Group., *Simple heuristics that make us smart*. New York: Oxford University Press, 1999.
- [21] L. Cosmides and J. Tooby, "Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty," *Cognition*, vol. 58, pp. 1-73, 1996.
- [22] G. Gigerenzer and K. Fiedler, "Minds in environments: The potential of an ecological approach to cognition," *Manuscript submitted for publication*, forthcoming.
- [23] T. P. Weber, "News from the realm of the ideal free distribution," *Trends in Ecology & Evolution*, vol. 13, pp. 89-90, 1998.
- [24] D. W. Stephens and J. R. Krebs, *Foraging theory*. Princeton, N.J.: Princeton University Press, 1986.
- [25] D. G. C. Harper, "Competitive foraging in mallards: 'Ideal free' ducks," *Animal Behaviour*, vol. 30, pp. 575-584, 1982.
- [26] A. E. Lamb and J. G. Ollason, "Foraging wood-ants *Formica aquilonia* yarrow (hymenoptera: formicidae) tend to adopt the ideal free distribution," *Behavioural Processes*, vol. 28, pp. 189-198, 1993.
- [27] T. C. Grand, "Foraging site selection by juvenile coho salmon: ideal free distributions of unequal competitors," *Animal Behaviour*, vol. 53, pp. 185-196, 1997.
- [28] M. B. Sokolowski, F. Tonneau, and E. Freixa i Baque, "The ideal free distribution in humans: an experimental test," *Psychon Bull Rev*, vol. 6, pp. 157-61, 1999.
- [29] G. J. Madden, B. F. Peden, and T. Yamaguchi, "Human group choice: discrete-trial and free-operant tests of the ideal free distribution," *J Exp Anal Behav*, vol. 78, pp. 1-15, 2002.
- [30] P. W. Glimcher, *Decisions, uncertainty, and the brain : the science of neuroeconomics*. Cambridge, Mass. ; London: MIT Press, 2003.
- [31] C. R. Gallistel, "Deconstructing the law of effect," *Games and Economic Behavior*, vol. 52, pp. 410-423, 2005.
- [32] A. K. Seth, "The ecology of action selection: Insights from artificial life," *Phil. Trans. Roy. Soc. B*, vol. http://www.nsi.edu/users/seth/Papers/Seth_PTRS.pdf, in press.
- [33] C. Rode, L. Cosmides, W. Hell, and J. Tooby, "When and why do people avoid unknown probabilities in decisions under uncertainty? Testing some predictions from optimal foraging theory," *Cognition*, vol. 72, pp. 269-304, 1999.
- [34] B. Seth and M. T. Peter, "Made to Measure: Ecological Rationality in Structured Environments," *Minds and Machines*, vol. V9, pp. 497-541, 1999.
- [35] Y. Niv, D. Joel, I. Meilijson, and E. Ruppin, "Evolution of Reinforcement Learning in Uncertain Environments: A Simple Explanation for Complex Foraging Behaviors," *Adaptive Behavior*, vol. 10, pp. 5-24, 2002.
- [36] D. J. Anderson, "Optimal foraging and the traveling salesman," *Theoretical Population Biology*, vol. 24, pp. 145, 1983.
- [37] R. Menzel and M. Giurfa, "Cognitive architecture of a mini-brain: the honeybee," *Trends in Cognitive Sciences*, vol. 5, pp. 62-71, 2001.
- [38] P. R. Montague, P. Dayan, C. Person, and T. J. Sejnowski, "Bee foraging in uncertain environments using predictive hebbian learning," *Nature*, vol. 377, pp. 725-728, 1995.
- [39] T. Keasar, E. Rashkovich, D. Cohen, and A. Shmida, "Bees in two-armed bandit situations: foraging choices and possible decision mechanisms," *Behav. Ecol.*, vol. 13, pp. 757-765, 2002.
- [40] J. C. Gittins, *Multi-armed Bandit Allocation Indices*. New York: Wiley, 1989.
- [41] J. S. Banks and R. K. Sundaram, "Switching Costs and the Gittins Index," *Econometrica: Journal of the Econometric Society*, vol. 62, pp. 687-694, 1994.
- [42] R. K. Sundaram, "Generalized Bandit Problems," Working Paper, Stern School of Business 2003.
- [43] J. Tackseung, "A survey on the bandit problem with switching costs," *De Economist*, vol. V152, pp. 513-541, 2004.
- [44] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement Learning: A Survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237-285, 1996.
- [45] W. K. Estes, "Individual behavior in uncertain situations: An interpretation in terms of statistical association theory.," in *Decision processes*, R. M. Thrall, C. H. Coombs, and R. L. Davies, Eds. New York: Wiley, 1954, pp. 127-137.
- [46] R. J. Meyer and Y. Shi, "Sequential Choice under Ambiguity: Intuitive Solutions to the Armed-Bandit Problem," *Management Science*, vol. 41, pp. 817-834, 1995.
- [47] P. W. Glimcher, M. C. Dorris, and H. M. Bayer, "Physiological utility theory and the neuroeconomics of choice," *Games and Economic Behavior*, vol. 52, pp. 213, 2005.
- [48] G. Morris, A. Nevet, D. Arkadir, E. Vaadia, and H. Bergman, "Midbrain dopamine neurons encode decisions for future action," *Nat Neurosci*, vol. 9, pp. 1057-1063, 2006.
- [49] S. J. Gould and R. C. Lewontin, "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme," *Proc R Soc Lond B Biol Sci.*, vol. 205, pp. 581-98, 1979.
- [50] D. M. Egelman, C. Person, and P. R. Montague, "A computational role for dopamine delivery in human decision-making," *J Cogn Neurosci*, vol. 10, pp. 623-30, 1998.
- [51] P. W. Glimcher, M. C. Dorris, and H. M. Bayer, "Physiological utility theory and the neuroeconomics of choice," *Games Econ Behav*, vol. 52, pp. 213-256, 2005.

- [52] A. Perez-Urbe, "Using a time-delay actor-critic neural architecture with dopamine-like reinforcement signal for learning in autonomous robots," in *Emergent Neural Computational Architectures based on Neuroscience* S. Wermter, J. Austin, and D. Willshaw, Eds. Heidelberg, Germany: Springer-Verlag, 2001, pp. 522-533.
- [53] M. Hammer, "An identified neuron mediates the unconditioned stimulus in associative olfactory learning in honeybees," *Nature*, vol. 366, pp. 59-63, 1993.
- [54] W. Schultz, "Reward signaling by dopamine neurons," *Neuroscientist*, vol. 7, pp. 293-302, 2001.
- [55] W. Schultz, "Behavioral theories and the neurophysiology of reward," *Annu Rev Psychol*, vol. 57, pp. 87-115, 2006.
- [56] D. E. Goldberg, "Probability Matching, the Magnitude of Reinforcement, and Classifier System Bidding," *Machine Learning*, vol. V5, pp. 407-425, 1990.
- [57] T. D. Wickens, *Elementary signal detection theory*. Oxford ; New York: Oxford University Press, 2002.
- [58] R. Noë and P. Hammerstein, "Biological markets," vol. 10, pp. 336, 1995.
- [59] R. Noë, J. A. R. A. M. v. Hooff, and P. Hammerstein, *Economics in nature : social dilemmas, mate choice, and biological markets*. Cambridge, UK: New York, NY Cambridge University Press, 2001.
- [60] K. E. Stanovich, "Is probability matching smart? Associations between probabilistic choices and cognitive ability," *Mem Cognit*, vol. 31, pp. 243-51, 2003.
- [61] D. R. Shanks, R. J. Tunney, and J. D. McCarthy, "A re-examination of probability matching and rational choice," *Journal of Behavioral Decision Making*, vol. 15, pp. 233-250, 2002.