Statistical Reasoning Strategies in the Pursuit and Evasion Domain

Sevan G. Ficici and Jordan B. Pollack

DEMO Lab, Brandeis University, Waltham MA 02454, USA
www.demo.cs.brandeis.edu

Abstract. Isaacs' treatise on differential games was a break-through for the analysis of the pursuit-and-evasion (PE) domain within the context of strategies representable by differential equations. Current experimental work in Artificial Life steps outside of the formalism of differential games, but the formalism it steps *into* is yet to be identified. We introduce a formulation of PE that allows a formalism to be developed. Our game minimizes kinematic factors and instead emphasizes the *informational* aspect of the domain. We use information-theoretic tools to describe agent behavior and implement a pursuit strategy based on statistical decision making; evaders evolved against this pursuit strategy exhibit a wide range of sophisticated behavior that can be quantitatively described. Agent performance is related to these quantifiables.

1 Introduction: Body and Mind in Pursuit and Evasion

Researchers in the field of Artificial Life (ALife) frequently turn to the domain of pursuit-and-evasion (PE) to study the (co)evolution of complex agent behavior. PE is of particular interest because it provides a parsimonious framework of agent interaction as well as ethological interest. Isaacs' ground-breaking theory of differential games [15] shows that, for a great many forms of the PE game, optimal (minimax) player strategies can be derived analytically from knowledge of the players' abilities. The applicability of differential game theory is constrained by two requirements: 1) a finite set of *state variables* must completely capture the instantaneous description of the game (e.g., agent positions, velocities, accelerations, etc.) necessary to compute the future unfolding of the game—the game is one of perfect information; 2) agent strategies must be expressible as differential equations that operate on these state variables. These constraints make the theory (and its more modern versions, e.g., [8]) a powerful model of agent *kinematics* (e.g., speed, maneuverability, etc.), as kinematics are particularly amenable to differential analysis. Indeed, differential game theory was constructed to answer questions about agents that are distinguished primarily in their kinematic abilities.

Much ALife research, however, steps outside of the formal assumptions made by Isaacs' theory; evolvable sensory apparatus [5] and evolvable information-processing substrates for agent control, such as artificial recurrent neural networks, move the PE domain beyond the purely kinematic realm to include an *informational* one as well. Clearly, differential game theory remains an influential formalism in the design and analysis of experiment kinematics. But, what formalism has been adopted to address the informational aspects of current experimental work? The lack of a rigorous metric of agent behavior has been recognized [7]—how do we ascertain and describe the sophistication of an agent? Recently, information-theoretic tools have been used to measure and adjust environmental complexity to facilitate agent learning in various domains, including PE [11, 13]; information theory has also come into use for the measurement of agent behavior, for example in a discrete state space [23] and a "linear" form of the PE game [10]. Here, we will introduce the use of information theory to measure agent behavior and environmental complexity in a two-dimensional, continuous-space pursuer-evader domain.

The purpose of replacing (stateless) minimax strategies with evolvable agent controllers is to place the pursuit and evasion roles within a coevolutionary setting such that an arms race between ever more sophisticated evasion and pursuit strategies arises [16, 19, 24, 6, 7, 12, 26]. Unfortunately, the envisaged coevolutionary arms race towards complexity is yet to be substantially realized, or, at least, observed; a number of problematic issues surrounding coevolutionary techniques are known to exist, such as the Red Queen Effect [6], mediocre stable-states [22], and the instability of two-population coevolution [4, 2, 9].

If we want an arms race to complexity, we must consider the three components that are responsible for the behavior of the typical PE agent: its sensory apparatus, kinematics, and processing abilities. Which of these three components possess the potential when evolved to engender an arms race to complexity? How might asymmetries between the pursuer and evader in terms of some of these components (say, their sensory and kinematic abilities) diminish the usefulness of evolving others (say, their computational abilities)? We assert that the role of processing ability is yet to be appreciated in its own right—that it has been overshadowed by kinematic and sensory factors.

In this paper, we present a simple formulation of the two-dimensional pursuer-evader game that, by giving the players kinematic parity and a form of sensory asymmetry, requires them to model opponent behavior in order to achieve optimal performance (in the game-theoretic sense). The behavior of an agent determines the complexity of the statistical model required to adequately represent it. The success of an agent reflects the power of its representational and modeling abilities. In this sense, there exists the potential for an arms race in complexity.

We begin, however, by using simple evolution to evolve evaders against hand-built predictors that employ statistical tools to model evader behavior. The questions of interest are: what happens when the statistical model of the pursuer is poor? What kind of behavior must an evader exhibit to defeat statistical prediction of some finite power? Would such evasion behavior resemble our intuitive understanding of *protean behavior* [19]? Our results show that evaders are able to evolve behaviors of substantial sophistication when placed in opposition to a powerful pursuit strategy.

We start with a formal definition of the game and discuss some game-theoretic features. We review the information-theoretic tools used to measure behavior. Experimental results are then discussed. Finally, we point to future work.

2 A Game of Incomplete Information

2.1 Definition

Our game is played in the real-valued, unbounded (two-dimensional) plane. Both agents move with *simple motion*: each agent moves at a constant velocity and is able to instantaneously change its direction by any amount. In our game, both agents move at the same velocity of one body-length per time-step. While the agents are able to freely pick a direction, they can only do so at the beginning of each time-step; they must then commit to moving in their chosen directions for the duration of one time-step. The pursuer can see the evader's current location, but the evader is completely blind. Thus, the pursuer has the opportunity to respond to the behavior of the evader, while the evader's behavior is ballistic.

Each game begins with the pursuer and evader occupying the same location, the plane's origin. Because we are interested in statistical characterizations of agent behavior, each game must be long enough for a reasonably representative sample of behavior to be made with respect to the type of statistic we wish to collect. (In practice, the behaviors of our evolving evaders quickly achieve steady state—initial transients are either very short or unfold so slowly as to be indistinguishable from steady-state behavior for our purposes.)

Thus, the terminating condition occurs when the final time-step transpires, rather than when the evader gets tagged or reaches some "escape" distance. The evader's job is to maximize the *average distance* between itself and the pursuer over the course of the game. The pursuer's job is to maximize the number of *tags*, which is subtly more general than minimizing distance, as we will discuss. A tag occurs when the pursuer comes within one body-length of the evader.

2.2 Game Theoretic Features

This game is designed to require players to induce models of opponent behavior (the type of model is discussed below). Here we explore the game-theoretic features that create such conditions. To begin, we wish to eliminate kinematic asymmetries that may alone favor one agent over the other. For this reason, we give both players identical kinematic abilities: simple motion with equal velocities. The choice to make our evader blind appears to contradict the care taken to give the agents an equal footing. To understand this choice, let us consider what happens if the evader has the ability to see. Further, let us assume the predictor is strictly trying to minimize distance, such that the game is clearly *zero-sum*.

If the sightedness of both players is *common knowledge* [18, 14], i.e., known by both players, then this modified game clearly calls for a *minimax* strategy [15] to be used. The minimax strategy calls for the pursuer to move directly towards the evader and the evader to move directly away from the pursuer. Any divergence from this strategy by an agent gives an advantage to its opponent;

if both agents observe their optimal strategies, then the distance between them does not change. Given our kinematic constraints, minimax allows neither the evader to increase its distance from the pursuer, nor the pursuer to near the evader, let alone tag it—a very dull game, indeed. This is why agent kinematics are made asymmetrical in most experimental work; the possibility of "cognitive" asymmetry is thus marginalized.

We can now appreciate what happens if the evader is blind. Because the evader is unaware of the pursuer's location, it can not simply move away. Instead, the evader must employ a mixed strategy that picks moves probabilistically. If the pursuer knows this to be the case, through common knowledge, then the pursuer is best advised to adopt a *statistical decision process* [27] based upon observations of the evader's behavior. If the pursuer's goal is not only to decrease the distance between itself and the evader, but also to do so as quickly as possible (in order to maximize the number of tags), then pursuit based upon statistical observation will, on average, very likely outperform (and certainly do no worse than) the simple pursuit strategy that minimax specifies above, provided that the statistical observations are suitably accurate. If the evader knows, again through common knowledge, that the pursuer is using such a strategy, then it is no less motivated to behave probabilistically.

In his discussion of pursuit games with incomplete information, Isaacs [15] explains that the importance of accurate prediction diminishes as the distance between opponents increases; if the "probability cloud" of possible future evader locations is small compared to the distance between the evader and the pursuer, then the cloud is reasonably treated as a single point by the pursuer—obviating the need to predict and allowing the pursuer to revert to the simple strategy of moving directly towards the evader. On the other hand, if the two agents are very close, particularly if their regions of possible future locations overlap, accurate prediction is of utmost importance. With spatial proximity, the statistical nature of the game becomes more prominent. For this reason, we choose to begin our game with the pursuer and evader on top of one another.

3 Observation and Statistical Reasoning

Because the evader moves at a constant rate, and commits to some direction of travel for the duration of each time-step, the pursuer need only observe the sequence of direction choices that the evader makes; if the pursuer accurately models the evader's decision process, then effective pursuit is possible. Though the directions in which the agents move (and the locations they occupy) vary continuously, the pursuer's *sensing* of movement has limited resolution: each move of the evader is perceived by the pursuer as being in one of eight directions, as shown in Figure 1. Thus, the pursuer's statistics model the *symbol string* generated by the evader's behavior, and not the evader's behavior directly.

To predict the future path of the evader, the pursuer uses its statistical model to generate the symbol string it expects the evader to produce in the coming time-steps. The pursuer then uses this symbol string to project the expected trajectory relative to the evader's actual current location. The pursuer then acts according to its policy, detailed below. Even if the symbol string is predicted with complete accuracy, the projected trajectory only approximates the path the evader will actually take, due to the pursuer's limited ability to sense and represent heading. Nevertheless, the eight-symbol resolution is high enough for adequate approximation in practice.

In its capacity as model-builder, our hand-built pursuer uses techniques that we, as experimenters, can use to quantitatively describe an evader's behavior. Our analytic approach comes from information theory [25]. The model of an evader is built by computing n^{th}*-order statistics* from observations of behavior. The order of the statistic refers to the amount of conditioning applied when computing the probabilities of various symbols occurring. The higher the order of our statistics, the more we can refine our expectations of what will happen in the future based on what we have observed further in the past. For example, a 0-order statistic states the probability of a symbol occurring, $p(S)$, without conditioning—based strictly on the number of times it is observed; if 10% of the symbols observed are the symbol 'X' then the 0-order probability of 'X' is $p(X) = 0.1$. A 1-order statistic states the probability of a symbol conditioned with respect to the symbol that immediately precedes it; if the symbol 'Y' is followed by 'X' 60% of the time, then one of the 1-order probabilities of 'X' is $p(X \mid Y) = 0.6$. We now know more about when to expect the symbol 'X'.

The *entropy* of an observed sequence describes the over-all predictability of the sequence, given a particular order statistic; higher entropy values indicate less predictability. Certain symbol sequences require higher order statistics to be taken than other sequences to achieve the same level of predictability—the same entropy. The minimal amount of behavioral history, or conditioning, required to maximize predictability is an indication of a sequence's *complexity*. Of equal importance is the degree to which a sequence becomes less predictable as the order of the statistic is decreased. These two characterizations of a symbol sequence provide a quantifiable description of behavior.

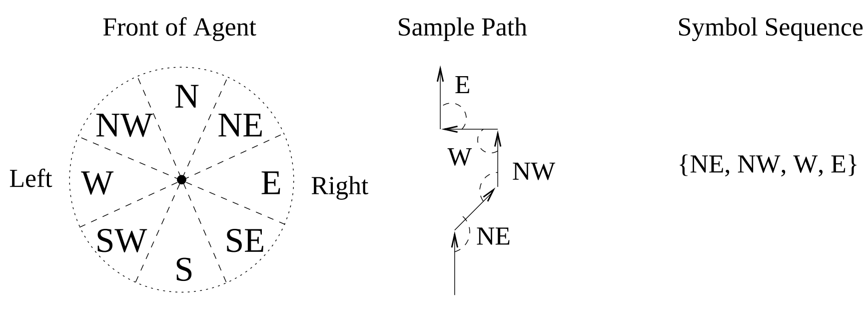

Fig. 1. Pursuer sensing of evader movement. Each turn of the evader generates one of eight symbols, depending on the angle of the turn relative to evader's previous heading.

4 Experiments

4.1 Setup

Evolving Evader Using an enhanced version of GNARL [3], we evolve artificial recurrent neural networks to control our evader agent. The GNARL algorithm adapts Evolutionary Programming (EP) techniques to the evolution of neural networks. Since the evader is blind, there are no inputs to the network; a single bias node, which has a constant activation, is used. The network has a single real-valued output that indicates the turn angle for the next move. Networks are allowed to have as many as 60 hidden units and 400 weights. These limits are not known to be optimal in any sense (casual examination of the larger evolved networks shows around 50 hidden nodes but only 100 weights). The initial population is comprised of random networks that have between 5 to 15 hidden units and 25 to 100 weights. The population size is 100. The fitness of a controller network is equal to the average distance it is able to maintain between the pursuer and the evader over the course of a game.

Hand-Built Pursuer The evader controller networks are evolved against a hand-built pursuit strategy that uses statistical modeling of evader behavior. This pursuit strategy is controlled by two parameters: 1) the power of the statistical model, which is specified by the order statistic, o, and 2) the number of time-steps into the future, t, that will be predicted by the model. The more the pursuit strategy looks into the future, the more opportunity it has to discover and exploit short-cuts to predicted future locations of the evader. But, with increased values of t comes increased risk of being misled by poor prediction.

At each time-step of the game, the pursuit strategy proceeds as follows: using the statistical model of order o, predict the evader's movement from its current location for t time-steps into the future; this yields the path, $\boldsymbol{p}$, of points $\boldsymbol{p} = \{p_1, p_2, ..., p_t\}$. Assuming the predictions to be correct, we know when the evader will arrive at each point, p_i, in the path, $\boldsymbol{p}$. If there exist any points to which the pursuer, by heading directly to them, can arrive before the evader, select the earliest such point in the path. Otherwise, move towards the first point on the path, p_1.

The pursuer computes its statistical model of evader behavior *tabula rasa* for each game. Because the model requires a suitably large sample of behavior to give meaningful predictions, we preface each match between the pursuer and an evader with a "warm-up" period of ten thousand time-steps during which the statistical model of the evader's ballistic behavior is induced. The length of this warm-up period represents a practical compromise between sample quality and simulator speed. After the model is built, the players are placed on the origin of the plane and the game proper begins, which lasts for one thousand time-steps.

4.2 Results

We outline above a statistical formalism with which we can characterize evader behavior, and describe how we embed this formalism into our hand-built pursuit strategy. We now evolve evaders against this strategy. By thus inserting our statistical formalism into the evolutionary process, however, we lose, ironically, the ability to use it as an objective measure of experimental results; while we do provide some quantitative results below, we must also pay careful attention to the qualitative results for meaningful understanding. Our formalism can more properly be used for agent analysis once both the evader and pursuer are made to coevolve; this is the subject of future work.

Pursuer Performance Figure 2 (left) captures a moment during a match. The square and the large circle represent the positions of the pursuer and evader at time t, respectively. The small circles are the pursuer's prediction (using 12^{th}-order statistics), made at time $t-1$, of the evader's path for the next twelve time steps. We see that the evader is exactly where the pursuer expected it to be at time t. The large 'X' represents the point on the evader's path where the pursuer expects to most quickly intercept the evader; it is towards this point that the pursuer moved directly from time-step $t-1$ to time t. Of course, interception will occur only if the predictions are correct. The evader's behavior might instead lead the pursuer into a false expectation of future locations—the only method by which the evader can open space between itself and the pursuer; this is what we see in Figure 2 (right), which shows a much weaker pursuer (using 3^{rd}-order statistics) playing against the same evader at the identical point in the match. Clearly, and to its detriment, this pursuer has a much different model of the evader and very different expectations of future evader behavior.

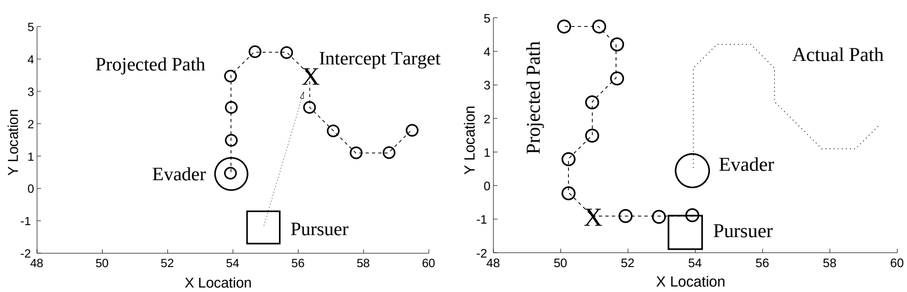

Fig. 2. Pursuit using 12^{th}-order statistics (left) and 3^{rd}-order statistics (right).

Evader Behavior Though the evaders are blind, they too must model their opponent to succeed. In this case, the model is induced through evolution and reflects the settings of the pursuer's two parameters: the order of pursuer's statistical observations, and the number of time-steps predicted into the future. A very wide variety of evader behaviors evolve in our experiments. The trajectories of two different evaders, A and B, are shown in Figure 3 (top left and right, respectively). Both were evolved in the same run with pursuit parameters set to $t = 37$ and $o = 4$; evader A is the best of generation 825, while B arrives much earlier in the run and is the best of generation 85. Evader A has a highly irregular, though smooth, trajectory that contains tight loops, broad and narrow turns, and some relatively straight stretches. Evader B's behavior is extremely regular (in fact periodic) and jagged, and displays two frequencies of movement (one superimposed on the other), which gives it a hint of self-similarity. Both evaders clearly maintain non-trivial internal state.

Below each trajectory, we graph the entropy of the trajectory (dotted line), the average distance maintained between the evader and the pursuer (solid line), and the average tag rate (dash-dot line); these values are given for a pursuer with parameter $t = 37$ and parameter o ranging from 0 to 9. Though the entropy of evader B, when measured with 0-order statistics, is actually higher than that of evader A (approximately 1.9 for B and 1.3 for A out of a maximum possible entropy of 3.0), the entropy of evader A falls only slightly as the order statistic is increased, whereas the entropy of evader B drops dramatically. These entropy curves reflect what is visibly obvious: evader A is very difficult to predict and evader B is very easy. Particularly, once the pursuer models evader B as a 6^{th}-order process, the regular behavior of the evader becomes evident: the observed entropy falls to a mere 0.16, the average distance the evader is able to maintain from the pursuer drops to 0.37 body-length units, and the pursuer's tag rate jumps to a very effective 0.98%. Changes in evader A's performance are evident as modeling power is increased, as well, but are not nearly as dramatic.

These graphs suggest that proteanism—adaptively unpredictable behavior—may be envisioned as a continuum; a behavior may be protean relative to a weak statistical model but not to a more powerful one. While a high entropy value implies a behavior that is difficult to predict, it does not necessarily imply an effective evasion strategy; Isaacs recognizes that, in a game of incomplete information, a tension may exist between moves that create uncertainty and moves that open distance. Good evasion behavior consists of a balanced mix of these types of move. Thus, while evader B has a higher 0-order entropy than evader A, its behavior does not produce such a mix and poorly leverages the uncertainty it generates.

As a simple control experiment, we modify the pursuer to move directly towards the evader's current location (as in minimax); the appropriate minimax response, i.e., straight fleeing, is evolved by the evader. This provides further empirical evidence that evolution is sensitive to the pursuit strategy and creates evasion behaviors accordingly. Thus, effective evasion is not synonymous with complex evasion; the complexity of evasion behaviors we see in our main experiment reflect the sophistication of our hand-built pursuit strategy. In *co*evolutionary frameworks, an absence of evader "proteanism," as conceived here, indicates a similar lack of pursuer sophistication.

5 Conclusion

We introduce a formulation of the pursuer-evader game that emphasizes its informational component. We discuss statistical tools that enable rigorous analysis of evader behavior and, thereby, allow the construction of a powerful pursuit strategy. Evolution against this pursuit strategy results in evasion behaviors of considerable sophistication. Agent performance is relatable to quantifiable features of behavior.

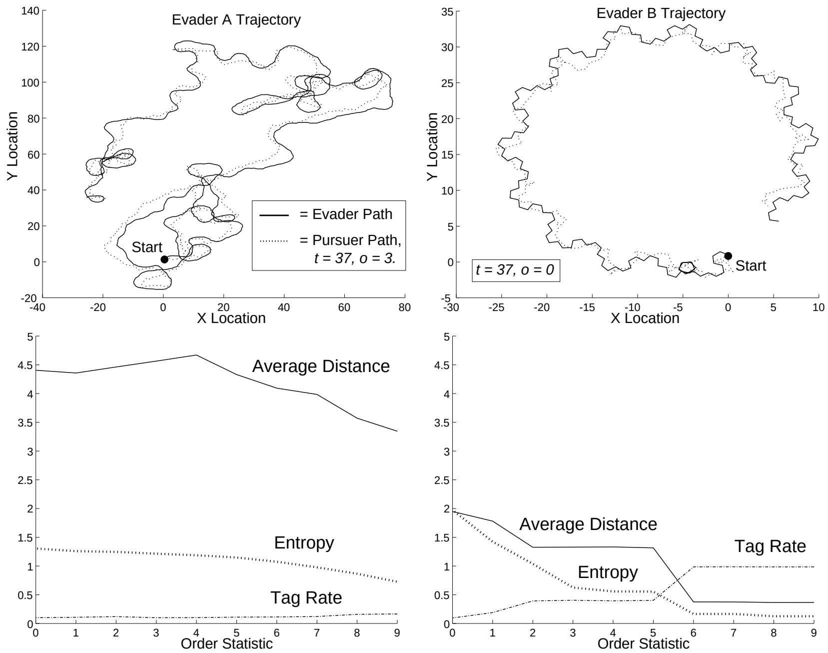

Fig. 3. Evader trajectories and performance curves. Trajectories begin at (0, 0).

Future work will replace the hand-built statistical mechanism of the pursuer with a coevolving substrate. The generality of our hand-built pursuer is what evokes evasion behavior with (often complex) statistical structure. But, unlike the blind evaders, pursuers can not operate ballistically—they must respond to observed behavior. If coevolving pursuers are not exposed to a wide enough diversity of evasion behavior, they will likely evolve pursuit strategies that lack generality; these specialized strategies will be suboptimal in the game-theoretic sense. Nevertheless, our statistical tools will be of particular use in tracking coevolutionary progress [9]. Finally, we will investigate sighted evaders.

6 Acknowledgments

The authors thank the members of DEMO for their insightful assistance, particularly Richard Watson, Ofer Melnik, and Hod Lipson.

References

1. C. Adami et al., editors. *Artificial Life VI (1998)*. MIT Press, 1998.
2. E. Akiyama and K. Kaneko. Evolution of communication and strategies in an iterated three-person game. In Langton and Shimohara [17], pages 150–158.
3. P. J. Angeline, G. M. Saunders, and J. B. Pollack. An evolutionary algorithm that constructs recurrent neural networks. *IEEE Transactions on Neural Networks*, 5:54–65, 1994.

4. S. Bullock. Co-evolutionary design: Implications for evolutionary robotics. In Moran et al. [20].
5. D. Cliff, P. Husbands, and I. Harvey. Evolving visually guided robots. In J.-A. Meyer, H. L. Roitblat, and S. W. Wilson, editors, *From Animals to Animats II (1992)*, pages 374–383. MIT Press, 1993.
6. D. Cliff and G. F. Miller. Tracking the red queen: Measurements of adaptive progress in co-evolutionary simulations. In Moran et al. [20], pages 200–218.
7. D. Cliff and G. F. Miller. Co-evolution of pursuit and evasion 2: Simulation methods and results. In P. Maes et al., editors, *From Animals to Animats IV*, pages 506–515. MIT Press, 1996.
8. R. J. Elliot and N. J. Kalton. The existence of value in differential games of pursuit and evasion. *Journal of Differential Equations*, 12:504–523, 1972.
9. S. G. Ficici and J. B. Pollack. Challenges in coevolutionary learning: Arms-race dynamics, open-endedness, and mediocre stable states. In Adami et al. [1], pages 238–247.
10. S. G. Ficici and J. B. Pollack. Coevolving communicative behavior in a linear pursuer-evader game. In Pfeifer et al. [21], pages 505–514.
11. J. A. Fletcher, M. A. Bedau, and M. Zwick. Effect of environmental structure on evolutionary adaptation. In Adami et al. [1], pages 189–198.
12. D. Floreano, S. Nolfi, and F. Mondada. Competitive co-evolutionary robotics: From theory to practice. In Pfeifer et al. [21], pages 515–524.
13. F. Gomez and R. Miikkulainen. Incremental evolution of complex general behavior. *Adaptive Behavior*, 5:317–342, 1997.
14. J. Harsanyi. Games with incomplete information played by 'bayesian' players. *Management Science*, 14:159–182,320–324,486–502, 1967.
15. R. Isaacs. *Differential Games.* John Wiley and Sons, New York, 1965.
16. J. Koza. *Genetic Programming.* MIT Press, 1992.
17. C. G. Langton and K. Shimohara, editors. *Artificial Life V (1996).* MIT Press, 1997.
18. D. Lewis. *Convention: A Philosophical Study.* Harvard University Press, Cambridge, Massachusetts, 1966.
19. G. F. Miller and D. Cliff. Protean behavior in dynamic games: Arguments for the co-evolution of pursuit-evasion tactics. In D. Cliff et al., editors, *From Animals to Animats III*, pages 411–420. MIT Press, 1994.
20. F. Moran et al., editors. *ECAL 3.* Springer, 1995.
21. R. Pfeifer et al., editors. *From Animals to Animats V.* MIT Press, 1998.
22. J. B. Pollack, A. Blair, and M. Land. Coevolution of a backgammon player. In Langton and Shimohara [17].
23. C. Pötter. Information-theoretic analysis of a mobile agent's learning in a discrete state space. In A. Eiben et al., editors, *PPSN V*, pages 521–529. Springer, 1998.
24. C. Reynolds. Competition, coevolution and the game of tag. In R. A. Brooks and P. Maes, editors, *Artificial Life IV*, pages 59–69. MIT Press, 1994.
25. C. E. Shannon and W. Weaver. *The mathematical theory of communication.* University of Illinois Press, Urbana, Illinois, 1963.
26. M. Wahde and M. G. Nordahl. Co-evolving pursuit-evasion strategies in open and confined regions. In Adami et al. [1], pages 472–476.
27. A. Wald. *Statistical Decision Functions.* Wiley, New York, 1950.

This article was processed using the LaTeX macro package with LLNCS style