Optimal and Non-Optimal Compromise Strategies in Action Selection

Frederick L. Crabbe
U.S. Naval Academy
Computer Science Department
572C Holloway Rd, Stop 9F
Annapolis, MD 21402
crabbe@usna.edu

Abstract

Devising a solution to the action selection problem rests on a definition of which actions are preferable over others. Among many properties suggested for action selection mechanisms, one prominent one is the ability to select compromise actions, i.e. actions that are not the best to satisfy any active goal in isolation, but rather compromise between the multiple goals. This paper performs a mathematical and experimental analysis of compromise behavior in an attempt to determine exactly how much compromise behavior aids an animat. It concludes that optimal compromise behavior is a small improvement over non-compromise behavior, but commonly proposed easy to compute compromise mechanisms are worse than non-compromise behavior. It further analyzes the properties of situations when compromise behavior is distinctly better than non-compromise behavior.

1. Introduction

Traditional Artificial Intelligence planning systems use search in order to fully characterize the space of actions a robotic agent (or animat) can select in a given situation. The animat considers the outcomes of possible actions into the future until it finds sequences of actions that achieve its goals. One feature of this approach is that given enough time, a planning system can determine the optimal actions for the animat. Of course, the issue of time is a fundamental problem for these planning systems: the animat may not have at its disposal the time needed in order to discover the optimal actions. In fact, often the amount of time required exceeds the age of the universe.

Behavior-based approaches to robotics and animats in general have been introduced to address these sorts of problems (Brooks, 1986, Arkin, 1998). These distributed reactive-style approaches are designed to generate "good enough" actions in a very small amount of time. Without optimality, there arises the important question of exactly what "good enough" means. In his now classic Ph.D. thesis, Tyrrell (Tyrrell, 1993) introduced a list of fourteen requirements for Action Selection Mechanisms. Of these, number twelve was "Compromise Candidates: the need to be able to choose actions that, while not the best choice for any one sub-problem alone, are best when all sub-problems are considered simultaneously." (p. 174) Tyrrell's list has had significant impact on the Action Selection field (e.g., Humphrys, 1996, Decugis and Ferber, 1998, Bryson, 2000, Girard et al., 2002), and a number of researchers have developed systems to meet the criteria he set out (e.g., Werner, 1994, Blumberg, 1994, Crabbe and Dyer, 1999). Meanwhile biologists have noted apparent compromise behavior among animals selecting between multiple targets (Morris et al., 1978, Latimer and Sippel, 1987, Bailey et al., 1990). The ability to consider compromise actions in an uncertain world makes great intuitive sense. When multiple goals interact, solving each optimally is not always optimal for the overall system. From the perspective of an animat in an environment where desired resources might go away at a moment's notice, a compromise movement in a direction in-between the resources might increase the animat's likelihood of obtaining one resource in the event that the other becomes unavailable.

Recent work has generated empirical results that seem to contradict the claim that the ability to consider compromise candidates is necessary (Bryson, 2000, Crabbe, 2002, Jones et al., 1999). Yet there have been few in-depth analyses of the nature of compromise actions and their effect on the overall success of an animat. This paper presents an extension of the work by Crabbe (Crabbe, 2002) and Hutchinson (Hutchinson, 1999) to investigate the nature of compromise actions, where an animat is faced with choosing between two destinations in the environment.

2. Problem Formulation

In this paper, we examine the situation where an animat has a goal to be co-located with one of two target locations in the environment. These could be, for instance, the locations of food, water, potential mates, or shelter. At any moment either or both of the targets can disappear from the environment. The animat must select an action that maximizes its chances of co-locating with a target before it disappears. This model is drawn from several scenarios in biology. For example, frogs or cricket males sometimes advertise for mates by emitting calls. The males may disappear with respect to the female through a cessation of signaling. This can occur either due to the actions of predators, the arrival of a competing female, or for internal reasons such as energy conservation. Another scenario in biology is that of a hunter such as a cat stalking prey such as birds in a flock, where an individual bird can fly at any moment (Hutchinson, 1999).

To approximate this type of scenario, we place an animat at the origin on a plane. Two targets are placed on the plane, one in the first and one in the second quadrant: T_a and T_b. The animat can sense the location of each target. Because the quality of the individual targets may vary, or the types of the targets may be different, the animat has two independent goals to be co-located with them. We call the strength of the goals G_a and G_b. The dynamism in the environment is represented with a probability p. This is the probability that any object in the environment will still exist after each time step. That is, any object will spontaneously disappear from the environment at each time step with probability $1-p$. Notationally, D_i^j is the distance from some location i to some location j, such that $D_{(0,0)}^{T_a}$ is the distance from the origin to the location of T_a. All distances are measured in the number of time steps it takes the animat to travel that distance.

When presented with this problem, the animat must select one action from the set of candidate actions. Because the task is a navigation task, all the candidate actions are movements to a location one time step away. Because the environment is continuous, not a discrete grid as in Tyrrell(Tyrrell, 1993), there are infinite candidate actions in our framework, each moving to a point on the unit circle centered at the animat's location. Comparing to Tyrrell's discrete world, our action choice may seem like some sort of "blending" of two choices. We argue that in fact, our formulation is identical to Tyrrell's except that the set of candidate actions our animats select from is much larger, i.e. infinite.

Once the animat has executed its action, it is again faced with the same action selection problem, unless one of the targets has disappeared and the best action is to move directly to the other target. As the animat repeatedly selects an action, the path it follows resembles a piece-wise linear approximation of a curved path (figure 1).

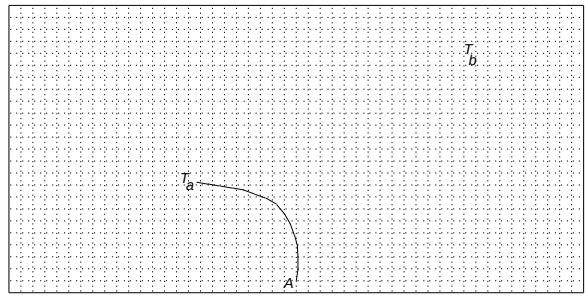

Figure 1: A possible path taken by an animat after repeated selection of a compromise candidate action. A marks the starting location of the animat, and T_a and T_b mark the locations of the targets.

It should be noted that the formulation used in this paper is one possible formulation. Another common formulation is where one of the goals is "proscriptive", that is, demands that certain actions not be performed. For example, a goal to avoid a hazard would proscribe against approaching the hazard. While Tyrrell explicitly used a proscriptive goal formulation as an example, he specifically hypothesized that "[m]ore generally, a *compromise candidate*, which might be beneficial to two or more systems to an intermediate degree, may be preferable to any of the candidates which are most beneficial for one system alone." (p. 170) (Tyrrell, 1993) This general hypothesis implies both the prescriptive version of the problem analyzed in this paper, as well as the proscriptive version[1]. Tyrrell's generalized hypothesis can be seen in the inclusion of prescriptive compromise behavior in many action selection mechanisms, such as Werner (Werner, 1994) and Montes-Gonzales et al. (Montes-Gonzales et al., 2000). Furthermore, outside the animat community, biologists have advocated in favor of the prescriptive version of compromise behavior for some time (Morris et al., 1978, Latimer and Sippel, 1987, Bailey et al., 1990).

3. Analytical Set-up

In order to investigate compromise candidates, we will analyze the initial configuration using Utility Theory (Howard, 1977). Utility Theory assigns a set of numerical values (utilities) to states of the world. These utilities represent the usefulness of that state to an animat. Expected Utility (EU) is a prediction of the eventual total utility an animat will receive if it takes a particular action in a particular state. The Expected Utility (EU) of taking an action A in a state S is the sum of the product

[1] By prescriptive, we mean that both goals are positive, enticing the animat to do something, rather than prohibiting actions.

of the probability of each outcome that could occur and the utility of that outcome:

$$EU(A|S) = \sum_{S_o \in Outcomes} P(S_o|A,S)U_h(S_o) \quad (1)$$

where $P(S_o|A,S)$ is the probability of outcome S_o occurring given that the animat takes action A in state S, and $U_h(S_o)$ is the historical utility of outcome S_o as defined below.

Assuming the animat is rational, the set of goals to consume objects will be order isomorphic[2] to the set of the animat's utilities of having consumed the objects. Therefore, EU calculated with utilities is order isomorphic with EU calculated with goals instead. For our purposes, we will assume that the goals and utilities are equivalent ($G_a = U(T_a)$).

Because a rational animat is expected to select the action with the largest EU, the historical utility of a state is the utility of the state plus future utility, or the max of the expected utility of the actions possible in each state:

$$U_h(S) = U(S) + \max_{A \in Actions} EU(A|S). \quad (2)$$

An animat can calculate EU using multiple actions in the future by recursively applying equations (1) and (2).

3.1 Optimal Behavior

We analyze compromise behavior by comparing a close approximation of optimal behavior with several non-optimal but easy to generate behaviors. We approximate the optimal behavior based on the technique used by Hutchinson (Hutchinson, 1999). This technique overlays a grid of points on top of the problem space and calculates the maximal expected utility of each location given optimal future actions. This is done recursively starting at the target locations and moving outward until stable values have been generated for all grid points.

The value we are trying to calculate is the expected utility of acting optimally at some location in a state where the two targets are still in the environment: $EU(Opt|T_a, T_b, (x,y))$. If θ is the angle of the optimal move for the agent at location (x,y) and (x',y') is 1 unit away from (x,y) in direction θ, then by equations 1 and 2 the expected utility of being at (x,y) is:

$$\begin{aligned} EU(Opt|T_a,T_b,(x,y)) =& p^2 EU(Opt|T_a,T_b,(x',y')) + \\ & p(1-p)EU(Opt|T_a,(x',y')) + \\ & p(1-p)EU(Opt|T_b,(x',y')), \end{aligned} \quad (3)$$

$$EU(Opt|T_a,(x',y')) = G_a p^{D^{T_a}_{(x',y')}}, \text{ and,} \quad (4)$$

$$EU(Opt|T_b,(x',y')) = G_b p^{D^{T_b}_{(x',y')}}. \quad (5)$$

the total expected utility is the expectation over four possible situations: both targets there, both targets gone, a there but b gone, and vice versa. When one of the targets disappears from the environment, the optimal action for the animat to take is to move directly to the other target, as shown in equations 4 and 5.

Using the above equations, the expected utility of each grid point in the environment can be calculated provided $EU(Opt|T_a, T_b, (x',y'))$ can be accurately determined and θ can be found. If (x',y') is a grid-point, this can be done easily, otherwise the local EU function must be interpolated from the expected utility values of the surrounding grid points. Using the interpolated surfaces for the local values of $EU(Opt|T_a, T_b, (x',y'))$, the value of θ can be determined by searching for the angle that maximizes the function described by equation 3. Once the expected utility is determined for a grid point, its value is then used to calculate the expected utility of its neighboring grid-points. This process is repeated until values are collected for all the grid points. Because the estimated utility value can change for a point when the values of its neighbors change, the values of all the points are repeatedly re-estimated until the values stabilize. The pseudocode for the algorithm is given in figure 2.

3.2 Action Selection Mechanisms

It is typically computationally prohibitive for an animat to calculate the optimal action using a technique similar to the one described in the previous section. Instead, most ethologists and AI researchers propose easy to compute action selection mechanisms that are intended to approximate the optimal behavior (Cannings and Orive, 1975, Fraenkel and Gunn, 1961, Römer, 1993). The mechanisms can be divided into two categories: those that select a single target and move directly toward it, and those that exhibit some sort of compromise behavior. In the former category, we will consider:

- Closest (C): select the closest target.
- Maximum Utility (MU): select the target with the higher utility (G_a or G_b).

[2] "Two totally ordered sets $(A, \leq)$ and $(B, \leq)$ are order isomorphic iff there is a bijection from A to B such that for all $a_1, a_2 \in A, a_1 \leq a_2$ iff $f(a_1) \leq f(a_2)$."(Weisstein, 2001)

```
Open is a list of grid-points that need to be updated.
Closed table of updated points.
N is the point currently being updated.
V_N is the current EU estimate at N.
repeat
    Open ← enqueue target locations
    Closed ← ∅
    repeat
      N ← dequeue from Open
      when N ∉ Closed
        EU(Opt|T_a, T_b, (x', y')) ← interpolated EU
                function from neighboring points
        V_N ← max equation 3
        Open ← enqueue neighbors of N
        Closed ← add N
    until Open = ∅
until convergence
```

Figure 2: Algorithm for estimating the expected utility at all the grid points in the environment.

- Maximum Expected Utility (MEU): select the target with the higher expected utility if it were the only target in the environment ($p^{D^{T_a}_{(0,0)}} G_a$ vs. $p^{D^{T_b}_{(0,0)}} G_b$)

Of action selection mechanisms that exhibit compromise behavior, we will examine:

- Forces (F): the animat behaves as if it has two forces acting on it, where the strength of the force is proportional to the utility of the target divided by the square of the distance between the animat and the target location (e.g. $M = \frac{G_a}{D^{T_a}_{(0,0)}}$).

- Signal Gradient (SG): The animat behaves as if it is following a signal gradient. The targets emit a simulated "odor" that falls with the square of the distance from the target. The initial strength of the odor is proportional to the utility of the target. The animat moves to the neighboring location that has the strongest odor as the sum of the odor emanating from each of the two targets.

- Exponentially Weakening Forces (EWF): The pulling effects of the targets in the previous two mechanisms fall quadratically with distance, whereas the expected utility of targets falls exponentially with distance. Exponentially Weakening Forces tries to model this by behaving just as with the Forces mechanism, but the strength of the forces falls exponentially with distance, rather than quadratically. The magnitudes of the two vectors are $G_a p^{D^{T_a}_{(x,y)}}$ and $G_b p^{D^{T_b}_{(x,y)}}$.

The expected utility of each of these mechanisms can be calculated for any particular scenario by using equations 3, 4 and 5. Let n be the number of time steps it takes to come within a single move of a target using one of the mechanisms. Let T_e be the the eventual target that the animat arrives at, and ϵ be the distance between the animat after n steps and T_e. Finally, let $D^{T_a}_x$ be the distance from the location of the animat after the xth time step to the target T_a; let V be the current strategy, and $EU(V|x)$ be the expected utility after time step x. For the first n time steps, the expected utility at any time step x of using strategy V can be figured by:

$$EU(V|x) = p^2 EU(V|x+1) + p(1-p)p^{D^{T_b}_{x+1}} G_b + p(1-p)p^{D^{T_a}_{x+1}} G_a, \quad (6)$$

where the location of $x+1$ is determined by the strategy. The expected utility of the final step (when the animat is within ϵ of T_e) is:

$$EU(V|n) = (1-p^\epsilon)p^\epsilon p^{D^{T_f}_{T_e}} G_f + (1-p^\epsilon)p^\epsilon G_e + p^{2\epsilon} G_e. \quad (7)$$

These two equations can be combined to yield:

$$EU(V|0) = p^{2n}\left[(1-p^\epsilon)p^\epsilon p^{D^{T_f}_{T_e}} G_f + (1-p^\epsilon)p^\epsilon G_e + p^{2\epsilon} G_e\right] + \sum_{i=1}^{n} p^{2i-1}(1-p)\left[p^{D^{T_b}_i} G_b + p^{D^{T_a}_i} G_a\right] \quad (8)$$

Using this equation, we can evaluate each of the action selection mechanisms by generating a large number of scenarios and comparing the expected utility of each of the mechanisms with each other and with optimal behavior.

4. Experiments

In the first set of experiments we calculate and compare the expected utility of various strategies. In the second set, we determine some properties of scenarios that lead to good behavior by some of the strategies. For the experiments, we generated 50,000 compromise scenarios. In these scenarios, Target a (T_a) was randomly placed in the x-range of $(-100; 0)$ and the y-range of $(0; 100)$. Target b (T_b) was randomly placed in the x-range of $(0; 100)$ and the y-range of $(0; 100)$. The utility of each target is randomly generated in the range of $(0; 100)$. We also randomly generated p, the probability that a target would remain in the environment per time step, in the range $[0.9; 1)$[3].

4.1 Strategy Comparison Experiments

For each scenario, we calculated the expected utility of each of the action selection mechanisms described in

[3]Below 0.9, the probability that the animat would reach either of the targets approaches zero and is therefore less interesting.

	C	MU	MEU
% best	0.1471	0.5309	99.32
avg % EU over C	0.0	4.130	12.62
avg % EU over MU	9.353	0.0	15.31
avg % EU over MEU	-5.960	-8.491	0.0

Figure 3: Comparison of non-compromise strategies. For each strategy, the percentage of situations where that strategy works best is provided, along with the average expected utility improvement of each strategy compared to the other strategies.

	F	SG	EWF	Optimal
avg	-4.074	-2.791	-2.465	1.121
best	4.844	4.819	20.56	22.73

Figure 4: Comparison of compromise strategies to the non-compromise MEU. For each strategy, the average expected utility improvement over MEU and the single best expected utility improvement over MEU is provided.

the previous section: closest (C), maximal utility (MU), maximum expected utility (MEU), forces (F), signal gradient (SG), and exponentially weakening forces (EWF). We also estimated the expected utility of optimal behavior using the technique in section 3.. We compare the three non-compromise mechanisms (C, MU, and MEU) in the table in figure 3. In the table, we show the percentage of scenarios each strategy was the best (had highest expected utility). We also compare the average percentage improvement of expected utility using each strategy to using the other strategies. We see that the maximum utility strategy is clearly the best of the three as it almost always selects the better target and its overall expected utility is from 12% to 15% better than the other two strategies. It is interesting to note that Closest is a better strategy than Maximum Utility. This is likely because the expected utility falls exponentially with distance, so that closer targets are more useful than better targets.

The table in figure 4 compares the compromise based mechanisms with MEU. It shows both the average % improvement over MEU and the % improvement over MEU in the single best scenario. There are three interesting things to note in this table. The first is that the optimal strategy is only 1.1% better than the non-compromise based MEU. This is somewhat surprising, as one would expect that the optimal behavior would be much higher than a non-compromise approach. This result is consistent with the non-continuous space experiments of Crabbe (Crabbe, 2002) and the much smaller study in Hutchinson (Hutchinson, 1999). The result might also explain the results of researchers who find compromise behavior unnecessary. If the environment is sufficiently forgiving, the small improvement of optimal compromise behavior may go unnoticed.

The second thing to note is that *all* of the non-optimal compromise based strategies are *worse* than the MEU strategy. In order to double check these surprising results, we selected 50 initial configurations for which we calculated the expected utility using each of the strategies and we re-estimated the expected utility using Monte Carlo simulation. For each situation we ran 100,000 trials where an animat ran the selected strategy, and at each time step we randomly determined if each target remained in the environment. A trial ended when a target was reached or both targets disappeared. The total utility of all reached targets was divided by the number of trials to estimate the expected utility. We found that of the 50 scenarios selected, the Monte Carlo estimated expected utilities agreed with the directly calculated expected utilities to, on average, four decimal places, lending support the the accuracy of figure 4. These results may further explain why some researchers have found that compromise behavior is unnecessary: the commonly used compromise strategies perform worse than a non-compromise strategy. With respect to natural animals, this result implies that animals that exhibit compromise behavior are either: a) behaving non-optimally; b) using an as yet unproposed compromise based action selection strategy (or at least one of which we are unaware); c) behaving in that manner for reasons other than purely to compromise between two targets (Hutchinson (Hutchinson, 1999) suggests several other possible reasons for what appears to be two-target based compromise behavior).

The final property of figure 4 to note is that EWF is the best performing of the easy to compute compromise strategies that we tested. While it is certainly not conclusive, this may imply that the approach of decreasing the influence of farther targets exponentially is a good one for developing action selection strategies. If we examine the *best* scenario for EWF, we see that it is nearly as good as the best scenario for optimal. This brings up the questions of *when* are optimal and EWF most beneficial.

4.2 Analysis Experiments

While the conclusions of the previous section may be true as an average across all the initial scenarios generated, there may be types cases where the Optimal or EWF strategies are significantly better than the MEU. In this section, we use the idea of information gain from information theory to attempt to determine if this is so. For example, in previous work on a discrete two-choice simplification of the problem, Crabbe (Crabbe, 2002) claimed that a $p > 0.9$ was most important for there to be good compromise behavior, followed by the distances to the targets to be equal.

4.2.1 Information Theory

In order to analyze the properties of the situations in which there are good compromise candidates, we use the idea of information gain. This is the same technique used in the decision tree algorithms ID3 and C4.5 (Mitchell, 1997, Quinlan, 1983, Quinlan, 1993). The technique begins with a data set and a classification of the data elements. Our data will be the situations generated in previous experiments, and our classification will be whether or not the scenario has a compromise strategy with an EU better than MEU. The technique then considers a set of n-ary attributes which partition the data set. For each attribute, the technique considers the information gain of applying that attribute to partition the data set. Information gain is the decrease in entropy of the data set, where entropy is:

$$E(S) = \sum_{i \in Categories} -p_i \log_2 p_i \tag{9}$$

S is a data set, i is a category of the classification, and p_i is the proportion of S categorized as i. Entropy measures the inverse of the purity of a data set with respect to the classification. A set is most pure when all the data have the same classification, and least pure when the data is classifies evenly across all categories.

The information gain of dividing set S based on attribute A is defined as:

$$Gain(S, A) = E(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} E(S_v), \tag{10}$$

where v is a value of an attribute, $|S|$ is a size of set S, S_v is the subset of S for which attribute A takes value v, The attribute with the largest information gain is the attribute which best partitions the data set, and in our case is the property of a situation which most determines if there is a good compromise candidate.

The data describing a situation is continuous information, but attributes are n-ary classifiers. To cope with this, the attributes partition the values into ranges, such as every 0.05. The attributes we use are listed below:

- $D^{T_a}_{(0,0)}, D^{T_b}_{(0,0)}, D^{T_b}_{T_a}$. These three attributes are the distances between the key locations in the situation. The distances can be anywhere from 0 to 224, and the attributes break them up into ranges of 5 units.
- G_a, G_b. The goal values from 0 to 100, broken into ranges of 5 units.
- p. The value of p from 0 to 1, broken into ranges of 0.005.
- $R(D^{T_a}_{(0,0)}, D^{T_b}_{(0,0)})$. This encodes the relationship between $D^{T_a}_{(0,0)}$ and $D^{T_b}_{(0,0)}$, while keeping the values between 0 and 1 using the formula $\frac{D^{T_a}_{(0,0)}}{D^{T_a}_{(0,0)}+D^{T_b}_{(0,0)}}$. It is broken into ranges of 0.05.
- $R(D^{T_a}_{(0,0)}, D^{T_b}_{T_a})$. This encodes the relationship between $D^{T_a}_{(0,0)}$ and $D^{T_b}_{T_a}$, while keeping the values between 0 and 1 using the formula $\frac{D^{T_a}_{(0,0)}}{D^{T_a}_{(0,0)}+D^{T_b}_{T_a}}$. It is broken into ranges of 0.05.
- $R(D^{T_b}_{(0,0)}, D^{T_b}_{T_a})$ This encodes the relationship between $D^{T_b}_{(0,0)}$ and $D^{T_b}_{T_a}$, while keeping the values between 0 and 1 using the formula $\frac{D^{T_b}_{(0,0)}}{D^{T_b}_{(0,0)}+D^{T_b}_{T_a}}$. It is broken into ranges of 0.05.
- $R(G_a, G_b)$. This encodes the relationship between G_a and G_b, while keeping the values between 0 and 1 using the formula $\frac{G_a}{G_a+G_b}$. It is broken into ranges of 0.05.
- $R_l(G_a, D^{T_a}_{(0,0)}, G_b, D^{T_b}_{(0,0)})$ This encodes a *linear* relationship between $G_a, D^{T_a}_{(0,0)}, G_b$ and $D^{T_b}_{(0,0)}$, while keeping the values between 0 and 1 using the formula $\frac{G_a D^{T_a}_{(0,0)}}{G_a D^{T_a}_{(0,0)}+G_b D^{T_b}_{(0,0)}}$. It is broken into ranges of 0.05.
- $R_e(G_a, D^{T_a}_{(0,0)}, G_b, D^{T_b}_{(0,0)})$ This encodes a *exponential* relationship between $G_a, D^{T_a}_{(0,0)}, G_b$ and $D^{T_b}_{(0,0)}$, while keeping the values between 0 and 1 using the formula $\frac{G_a p^{D^{T_a}_{(0,0)}}}{G_a p^{D^{T_a}_{(0,0)}}+G_b p^{D^{T_b}_{(0,0)}}}$. It is broken into ranges of 0.05.

4.2.2 Optimal

We first classify the data based on whether the performance of the optimal behavior is a 1% or greater improvement over the performance of MEU for that scenario. Using equation 10, we find that attribute with the largest information gain is $R_e(G_a, D^{T_a}_{(0,0)}, G_b, D^{T_b}_{(0,0)})$ with a gain of 0.07243. Figure 5 shows a breakdown of the data as partitioned by the attribute. The x-axis is the range of values of the attribute, and the y-axis is the percentage of the data that falls in that range and is positively classified (the performance of the optimal behavior is a 1% or greater improvement over the performance of MEU). In the figure we see that the greatest likelihood of good compromise behavior is when the attribute is close to 0.5. This means that if $G_a p^{D^{T_a}_{(0,0)}}$ is close to $G_b p^{D^{T_b}_{(0,0)}}$ (when the expected utilities of the two targets are roughly equal), it is more likely that optimal behavior will confer a large advantage. This makes sense that the good compromise behavior would occur when the targets have the same expected utility.

The attribute with the next largest information gain is $R(D^{T_a}_{(0,0)}, D^{T_b}_{(0,0)})$ with a gain of 0.06917. Figure 6 shows a breakdown of the data as classified by the attribute. Once again, we see that the greatest likelihood of good

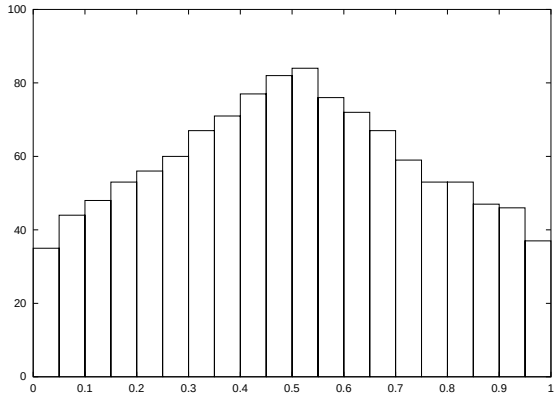

Figure 5: Breakdown of attribute $R_e(G_a, D^{T_a}_{(0,0)}, G_b, D^{T_b}_{(0,0)})$ when comparing optimal to MEU.

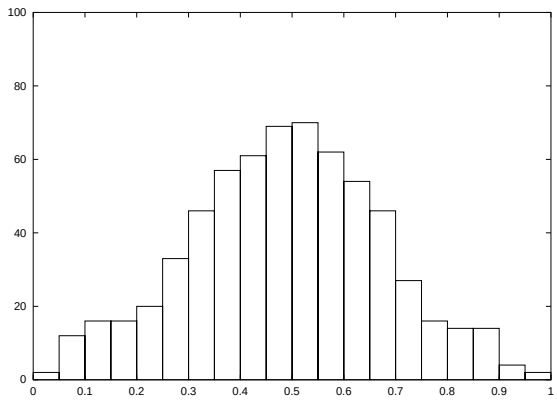

Figure 6: Breakdown of attribute $R(D^{T_a}_{(0,0)}, D^{T_b}_{(0,0)})$ when comparing optimal to MEU.

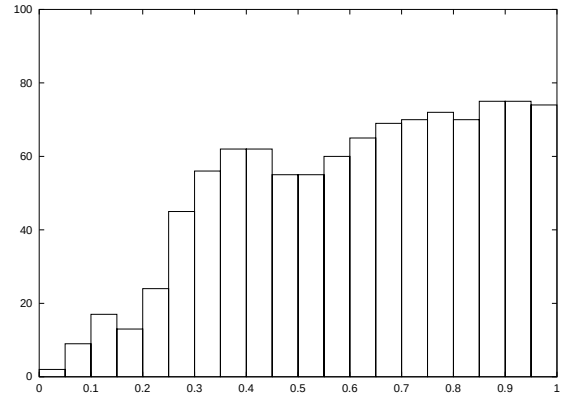

Figure 7: Breakdown of attribute $R(D^{T_a}_{(0,0)}, D^{T_b}_{T_a})$ when comparing optimal to MEU.

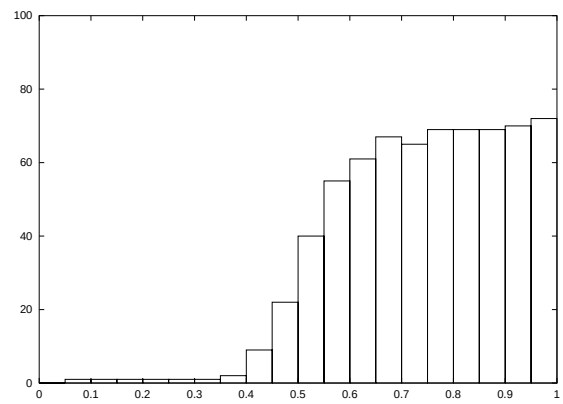

Figure 8: Breakdown of attribute $R(D^{T_a}_{(0,0)}, D^{T_b}_{T_a})$ when comparing EWF to MEU.

compromise behavior is when the attribute is close to 0.5, or in this case when $D^{T_a}_{(0,0)}$ and $D^{T_b}_{(0,0)}$ are roughly equal. We believe this is another manifestation of the need for a balance in expected utility for there to be good compromise behavior.

The next two attributes with high gain are $R(D^{T_a}_{(0,0)}, D^{T_b}_{T_a})$ and $R(D^{T_b}_{(0,0)}, D^{T_b}_{T_a})$, both with a gain of 0.027. Figure 6 shows a breakdown of the data as classified by $R(D^{T_a}_{(0,0)}, D^{T_b}_{T_a})$ (the plot of the other attribute is qualitatively identical). While the data in this figure is less distinct than in the previous figures, it does appear to show that there are *few* good compromises when the targets are relatively close to the animat but far away from each other.

4.2.3 Exponentially Weakening Forces

Although on average EWF performs worse than MEU, it is interesting to analyze the situations in which it performs better. If these situations could be recognized, then EWF could be applied only when they occur. Furthermore, it may lead to insight on why EWF outperformed the other non-optimal strategies. The data is classified based on whether or not EWF out performs MEU. The top two attributes in terms of information gain are $R(D^{T_a}_{(0,0)}, D^{T_b}_{T_a})$ and $R(D^{T_b}_{(0,0)}, D^{T_b}_{T_a})$, with gains of 0.1620 and 0.1569 respectively. Figure 8 shows a breakdown of the data as classified by $R(D^{T_a}_{(0,0)}, D^{T_b}_{T_a})$ and figure 9 shows a breakdown of the data as classified by $R(D^{T_b}_{(0,0)}, D^{T_b}_{T_a})$.

These graphs strongly suggest that good compromises using EWF occur when the targets are relatively far away from the origin and close to each other. This is backed up by examining the next highest attribute, $D^{T_b}_{T_a}$ (0.09223) in figure 10. Here we see that good compromises are more likely when $D^{T_b}_{T_a}$ is small, that is, the two targets are close together.

The last attribute we examine is p, with an information gain of 0.07903. Figure 11 shows its breakdown. It shows that good performance peaks at a p range of 0.91-

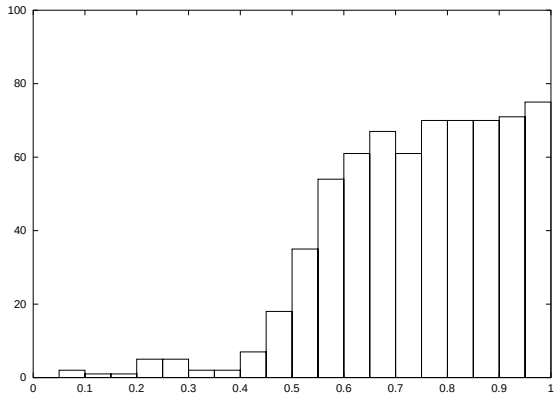

Figure 9: Breakdown of attribute $R(D^{T_b}_{(0,0)}, D^{T_b}_{T_a})$ when comparing EWF to MEU.

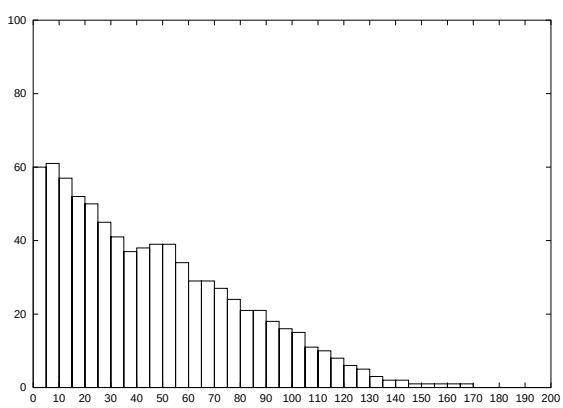

Figure 10: Breakdown of attribute $D^{T_b}_{T_a}$ when comparing EWF to MEU.

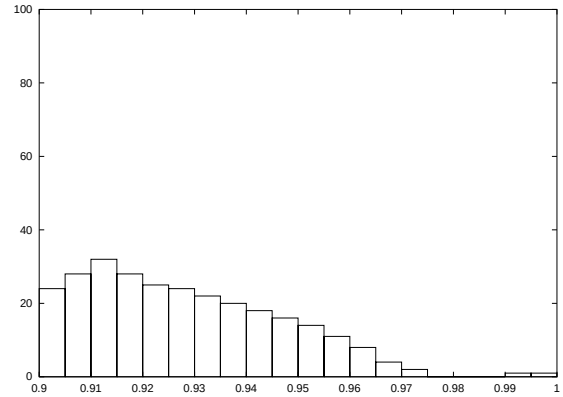

Figure 11: Breakdown of attribute p when comparing EWF to MEU.

0.915. This is consistent with Crabbe's previous results that good compromises occur with $p > 0.9$. That study only looked at p with a granularity of 0.1, and was thus unable to determine which values of $p > 0.9$ were best.

There is some overlap between the good attributes for the optimal behavior and for EWF, namely that the targets be relatively close together and far away from the animat. We have no evidence to explain why $R_e(G_a, D^{T_a}_{(0,0)}, G_b, D^{T_b}_{(0,0)})$ does not appear to be a factor for EWF, instead having a paltry information gain of 0.004855.

5. Conclusions and Future Work

In this paper we have analyzed the properties of action selection mechanisms in a scenario that has been of interest to both biologists and AI researchers in the past. We believe the paper makes several contributions. First, it derives a closed form equation for calculating the expected utility of any strategy in the multiple goal scenario (equation 8). This equation can also be trivially extended to any number of goals. As part of ongoing work, it has been modified to describe a proscriptive goal scenario was well. Next, the paper shows that among non-compromise strategies, closest is better than maximum utility, but the maximum expected utility is the best of the three. Maximum expected utility should be used if a non-compromise strategy is desired. When comparing compromise strategies, this paper concludes that optimal behavior performs only slightly better than maximum expected utility, and all other known compromise strategies perform worse than maximum expected utility. From this we conclude that animals that exhibit apparent compromise behavior are either using some unknown strategy or are doing so for some other reason. Of the non-optimal compromise strategies tested, exponentially weakening forces was the best, suggesting that if a better non-optimal compromise strategy exists, it may

also involve forces that fall exponentially with distance. This is a clear area of potential future work. When analyzing when the optimal and exponentially weakening forces strategies are better, we concluded that good optimal strategies occur when the expected utility of the two targets are equal. We further concluded that good scenarios for the optimal strategy occur when the targets are far away from the origin and close together. For the exponentially weakening forces strategy, we also concluded that good scenarios occur when the targets are far away and close together. Explaining why equality of the expected utility of the targets does not appear to be a factor is another area of future work.

Acknowledgements

We would like to thank Chris Brown and Rebecca Hwa for many wonderful discussions. We would also like to thank the anonymous reviewers for several helpful comments. This work was sponsored in part by a grant from the Office of Naval Research, number N0001403WR20241.

References

Arkin, R. (1998). *Behavior-Based Robotics*. MIT Press, Cambridge, MA.

Bailey, W. J., Cunningham, R. J., and Lebel, L. (1990). Song power, spectral distribution and female phonotaxis in the bushcricket requena verticalis (tettigoniiddae: Orthoptera): active femail choice or passive attraction. *Animal Behavior*.

Blumberg, B. M. (1994). Action-selection in hamsterdam: Lessons from ethology. In *Proceedings of the Third International Conference on Simulation of Adaptive Behavior*.

Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2:14–23.

Bryson, J. (2000). Hierarchy and sequence vs. full parallelism in action selection. In *Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*.

Cannings, C. and Orive, L. M. C. (1975). On the adjustment of the sex ratio and the gregarious behavior of animal populations. *Journal of Theoretical Biology*, 55:115–136.

Crabbe, F. (2002). Compromise candidates in positive goal scenarios. In *Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*.

Crabbe, F. L. and Dyer, M. G. (1999). Second-order networks for wall-building agents. In *Proceedings of the International Joint Conference on Neural Networks*.

Decugis, V. and Ferber, J. (1998). An extension of maes' action selection mechanism for animats. In *Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*.

Fraenkel, G. S. and Gunn, D. L. (1961). *The Orientation of Animals*. Dover, New York.

Girard, B., Cuzin, V., Guillot, A., Gurney, K. N., and Prescott, T. J. (2002). Comparing a brain-inspired robot action selection mechanism with 'winner-takes-all'. In *Proceedings of the seventh international conference on simulation of adaptive behavior: From animals to animats*, pages 75–84.

Howard, R. (1977). Risk preference. In *Readings in Decision Analysis*. SRI International, Menlo Park, CA.

Humphrys, M. (1996). *Action Selection methods using Reinforcement Learning*. PhD thesis, University of Cambridge.

Hutchinson, J. (1999). Bet-hedging when targets may disappear: optimal mate-seeking or prey-catching trajectories and the stability of leks and herds. *Journal of Theoretical Biology*, 196:33–49.

Jones, R. M., Laird, J. E., Nielsen, P. E., Coulter, K. J., Kenny, P., and Koss, F. V. (1999). Automated intelligent pilots for combat flight simulation. *AI Magazine*, 20(1).

Latimer, W. and Sippel, M. (1987). Acoustic cues for female choice and male competition in the tettigonia cantans. *Animal Behavior*.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, Boston.

Montes-Gonzales, F., Prescott, T. J., Gurney, K., Humphrys, M., and Redgrave, P. (2000). An embodied model of action selection mechanisms in the vertebrate brain. In *Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, pages 157–166.

Morris, G. K., Kerr, G. E., and Fullard, J. H. (1978). Phonotactic preferences of female meadow katydids. *Canadian Journal of Zoololgy*.

Quinlan, J. R. (1983). Learning efficient classification proceedures and their application to chess and games. In Michalski, R., Carbonell, J., and Mitchell,

T. M., (Eds.), *Machine Learning: an artificial intelligence approach*. Morgan Kaufmann, San Mateo, CA.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.

Römer, H. (1993). Environmental and biological constraints for the evolution of long-range signalling and hearing in acoustic insects. *Philosophical Transactions of the Royal Society B*, 340:179–185.

Tyrrell, T. (1993). *Computational Mechanism for Action Selection*. PhD thesis, University of Edinburgh.

Weisstein, E. W. (2001). *Eric Weisstein's World of Mathematics*. http://mathworld.wolfram.com/.

Werner, G. M. (1994). Using second order neural connection for motivation of behavioral choices. In *Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, pages 154–161.