On Compromise Strategies for Action Selection with Proscriptive goals

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

Abstract

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 an analysis of compromise actions in situations where the agent has one proscriptive goal. It concludes that optimal compromise behavior looks quite different from what was expected, and, while optimal compromise actions are beneficial to an agent, the benefit is often small compared to greedy algorithms. It goes on to suggest that much of the discussion about compromise behavior is the result of an equivocation on its definition, and it proposes a new compromise behavior hypothesis.

1 Introduction

Traditional Artificial Intelligence planning systems use search in order to fully characterize the space of actions a robotic agent can select in a given situation. The agent 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 action sequence for the agent. Of course, the issue of time is a fundamental problem for these planning systems: the agent 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 agents 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 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." [Tyrrell, 1993, p. 174] Tyrrell's list has had significant impact on the Action Selection field [Humphrys, 1996; Decugis and Ferber, 1998; Bryson, 2000; Girard *et al.*, 2002, e.g.], and a number of researchers have developed systems to meet the criteria he set out [Werner, 1994; Blumberg, 1994; Crabbe and Dyer, 1999; Avila-Garcia and Canamero, 2004, e.g.]. Meanwhile biologists and ethologists have noted apparent compromise among animals in several scenarios.

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. Yet, recent work has generated empirical results that seem to contradict the claim that the ability to consider compromise candidates is necessary [Jones *et al.*, 1999; Bryson, 2000; Crabbe, 2004]. Despite this, there have been few in-depth analyses of the nature of compromise actions and their effect on the overall success of an agent. This paper presents an extension of the work by Hutchinson [1999] and Crabbe [2004] to investigate the nature of compromise actions in various environmental conditions, concluding that: optimal compromise behavior is qualitatively different from what might be expected; optimal compromise behavior provides less benefit than expected in the scenarios tested; and the apparent disagreement about the utility of compromise behavior might possibly arise from an equivocation on its definition.

2 Problem Formulation

The action selection problem we will discuss in this paper depends on the types of actions the agent can select, the types of goals the agent pursues, and the formal representation of the problem.

2.1 Actions

When designing an action selection system, the character of the "actions" selected by the agent affect the behavior exhibited. For instance, there is a clear difference between an agent selecting the action *contract left quadricep 3 cm*. and the action *go to the refrigerator*. The distinction is based on the level of specificity given by the action; the former is as specific as possible, while the latter leaves much room for interpretation on how it is to be accomplished. In this paper we will define our domain to be that of navigation of a mobile agent, similar to several authors' simulated domains [Maes, 1990; Tyrrell, 1993] or navigating mobile robots [Choset *et al.*, 2005]. The space will be continuous, but time will be discrete, such that the action at each time step is defined as a movement 1 distance unit at any angle. The importance of this choice of the definition of "action" will be discussed in Section 6.

2.2 Goals

Tyrrell famously defined compromise as follows: "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." [1993, p. 170] The problem of compromise in action selection has multiple guises. One fundamental distinction pivots on the nature of the involved goals: are they prescriptive or proscriptive? Prescriptive goals encourage an agent to take some action or sequence of actions in order to be satisfied. These goals are typically satisfied by a final consumatory act. Proscriptive goals encourage an agent to *not* perform certain actions in certain situations. These goals are typically not satisfied by a particular action, but can be said to have been satisfied over a period of time if offending actions are not performed. The nature of a possible compromise scenario changes depending on whether there are two prescriptive goals or one prescriptive and one proscriptive goal.

Two Prescriptive Goals

In a two-prescriptive-goal case, an agent has goals 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 agent 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]. Several action selection mechanisms, such as Werner [1994] and Montes-Gonzales et al. [2000] have been specifically designed to exhibit this sort of compromise. Further, biologists and ethologists have advocated in favor of the prescriptive version for some time [Morris *et al.*, 1978; Lorenz, 1981; Latimer and Sippel, 1987; Bailey *et al.*, 1990].

Crabbe [2004] gave strong evidence that while this sort of behavior is seen in nature, it confers little absolute advantage to the agent. In particular: the optimal compromise strategy performed only slightly better than the best non-compromise (or greedy) strategy and all other known compromise strategies perform worse than maximum expected utility. They conclude that "animals that exhibit apparent compromise [in the 2 prescriptive goal case] are either using some unknown strategy or are doing so for some other reason." [Crabbe, 2004] This paper discusses the implications and a possible explanation of Crabbe's result in greater detail in Section 6 below.

One Prescriptive, One Proscriptive Goal

Although the two-prescriptive-goal scenario has had significant impact on the action selection community, the more famous of the two compromise scenarios discussed here is when there is one prescriptive goal and one proscriptive goal.

> "...proscriptive sub-problems such as avoiding hazards should place a demand on the animal's actions that it does not approach the hazard, rather than positively prescribing any particular action. It is obviously preferable to combine this demand with a preference to head toward food, if the two don't clash, rather than to head diametrically away from the hazard because the only system being considered is that of avoid hazard" [Tyrrell, 1993]

As the quote above indicates, the idea that compromise actions are especially beneficial in the proscriptive goal case is intuitively appealing. Further, examples of this appear in the ethological literature. Blue herons will select sub-optimal feeding patches to avoid predation by hawks in years when the hawk attacks are frequent [Caldwell, 1986]. Similar behavior has been shown in sparrows [Grubb and Greenwald, 1982], minnows [Fraser and Cerri, 1982], pike and sticklebacks [Milinksi, 1986]. At the motor level, geese and other *anatidae* who are offered food by a human can sometimes exhibit behavior where the neck muscles for both a feeding behavior and a recoiling behavior are activated, causing a trembling in the neck [Lorenz, 1981].

The purpose of this paper is to provide an analysis of the proscriptive goal scenario using the techniques developed by Crabbe for the prescriptive goal scenario, to determine both the amount of benefit of compromise actions, as well as under what conditions compromise actions are the most useful.

2.3 Formal Model

To approximate the scenario described in the quote above, we examine a continuous environment with a target t and a danger d, corresponding to a resource such as a mate and a predator respectively. At any time the target can disappear from the environment (e.g. the prospective mate stops signaling) with a probability $1-p_t$, and the danger can disappear (e.g. the predator becomes bored and wanders off) with a probability $1-p_d$. That is, at each time step, the target remains in the environment with probability p_t and the danger remains in the environment with probability p_d. Also at each time step, there is a probability $p_n(d)$ that the predator will *not* strike or pounce on the agent. This probability is a function of the distance between the agent and the danger. The experiments in this paper use four different functions to generate the $p_n(d)$. The agent also has a goal level associated with the target and the danger, (G_t and G_d) that can vary with the quality of the resource and the damage due to the predator. Notationally, $\overline{i,j}$ is the distance from some location i to some location j. All distances are measured in the number of time steps it takes the agent to travel that distance.

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 agent. Expected Utility (EU) is a prediction of the eventual total utility an agent 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 of the probability of each outcome that could occur and the utility of that outcome:

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

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

Assuming the agent is rational, the set of goals to consume objects will be order isomorphic[1] to the set of the agent'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 ($U(t) = G_t$).

Because a rational agent 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:

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

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

3.1 Optimal Behavior

We analyze compromise 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 dynamic programming technique adapted by Crabbe from 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 λ in a state where the target and the danger are still in the environment: $EU(O|t, d, \lambda)$. If θ is the angle of the optimal move for the agent at location λ and λ' is 1 unit away from λ in direction θ, then by equations 1 and 2 the expected utility of being at λ is:

$$\begin{aligned} EU(O|t,d,\lambda) =& p_t p_d p_n(\lambda) EU(O|t,d,\lambda') + \\ & p_t(1-p_d)EU(O|t,\lambda') + \\ & p_d(1-p_n(\lambda))G_d, \\ & (1-p_t)p_d p_n(\lambda)EU(O|d,\lambda') \end{aligned} \tag{3}$$

$$EU(O|t,\lambda) = G_t p^{\overline{\lambda,t}}, \text{ and,} \tag{4}$$

$$\begin{aligned} EU(O|d,\lambda) =& p_n(\lambda')p_d EU(O|d,\lambda') + \\ & (1-p_n(\lambda'))G_d. \end{aligned} \tag{5}$$

The total expected utility is the expectation over four possible situations: both target and danger are still there, but the danger does not strike; the target remains, but the danger disappears; the danger remains and strikes the agent; and the target disappears, the danger remains but the danger does not strike. When only the target remains, the optimal strategy is to go straight to the target, as in equation (4). When the target disappears but the danger remains, the agent must flee to a safe distance from the danger, as in equation (5). A safe distance is a variable parameter called the danger radius. Once the agent is outside the danger radius, it presumes that it is safe from the danger.

Using the above equations, the expected utility of each grid point in the environment can be calculated provided $EU(O|t, d, \lambda')$ can be accurately determined and θ can be found. Since λ' is most likely between grid-points, 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 $EU(O|t, d, \lambda)$, 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 1.

```
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(O|t,d,λ) ← interpolated EU
                    function from V_N at neighboring points
            V_N ← max equation 3
            Open ← enqueue neighbors of N
            Closed ← add N
    until Open = ∅
until convergence
```

Figure 1: The dynamic programming algorithm for estimating the expected utility at all the grid points in the environment.

3.2 Other Action Selection Mechanisms

It is typically computationally prohibitive for an agent to calculate the optimal action using a technique similar to the one described in the previous section. Instead, many 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]. In addition to the optimal strategy described above, we also examine three other action selection strategies:

- Direct: The agent moves directly to the target, ignoring the danger. This is a non-compromise strategy that one would expect to do poorly.
- Max goal: This strategy moves directly to the target unless the agent is within the danger zone. Within the danger zone, the agent moves directly away from the danger until it leaves the danger zone. This strategy zig-zags along the edge of the danger zone as the agent moves toward the target. Max Goal is also a greedy strategy that only acts upon one goal at a time.
- Skirt: This strategy moves directly toward the target unless such a move would enter the danger zone. In this case, the agent moves along the edge of the danger zone until it can resume heading directly to the target. Skirt is also primarily a greedy strategy. Outside the danger radius, the agent moves straight to the target. Inside the danger radius the agent moves straight away from the danger. At the edge of

[1] "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]

the danger radius the behavior is still optimal for the avoid danger goal, as any movement not into the danger zone is equally optimal. With respect to the target goal, the movement is sub-optimal.

The expected utility of each of these mechanisms can be calculated for any particular scenario by using equations 3, 4 and 5, where the action θ is the one recommended by the strategy, not the optimal action.

4 Experiments

The experiments were designed to determine how much better the optimal strategy is over the other strategies, as well as qualitatively examine what sorts of compromise actions are exhibited by the optimal strategy. In all of the trials, a target was placed at $(50, 90)$ with a $G_t = 100$ and a danger was placed at $(60, 50)$ with $G_d = -100$. In each trial, a p_t was selected in the range $[0.95; 1)$, and p_d was selected in the range $[0.5; 1)$. The $p_n(d)$ function was one of four functions, all of with with a danger radius of 20:

- Linear A: $p_n(d) = 0.04d + 0.2$ when $d \leq 20$, 1 otherwise.
- Linear B: $p_n(d) = 0.005d + .9$ when $d \leq 20$, 1 otherwise.
- Exponential: $p_n(d) = d^2/400$ when $d \leq 20$, 1 otherwise.
- Sigmoid: $p_n(d) = 1/(1 + 1.8^{10-d})$ everywhere.

Linear A was selected as a baseline strategy where the probability of a strike was high near the danger, but low at the edge of the danger zone. Linear B was selected to make the chance of a strike low overall, thus increasing the tendency to stay in the danger zone longer, for more compromise actions. Exponential has a high probability of a strike for much of the danger zone, but drops off sharply at the edge, perhaps encouraging compromise behavior near the edge. Sigmoid should resemble exponential, but the area with low strike probability is larger, and there is the possibility of some strike for every location in the environment, not just inside the danger radius.

Once the scenario was generated, the expected utility for each of the three non-optimal strategies and the optimal strategy was calculated for 200 points in the environment.

5 Results

Figure 2 shows the results of the optimal strategy when $p_t = 0.995, p_d = 0.99$, and $p_n(d)$ is Linear A. There are two interesting properties to note. First, within the danger zone, there is little display of compromise action; the agent flees directly away from the danger at all locations, ignoring the target. Second, there *is* compromise action displayed outside the danger zone, to the lower right. While this makes sense (the agent would want to take the shortest path around the danger zone) it does not fit into the common conception of compromise action. In that area of the environment, the goal to avoid the danger would not be active (since the agent is too far away from the danger) and thus one would expect it to have no effect of the action selected.

Figure 3 shows the optimal strategy when $p_t = 0.995, p_d = 0.5, p_n(d)$ is Linear A. The main difference is that the compromise action in the lower right is less pronounced. The optimal strategy is to assume that the danger will disappear by the time the agent gets there. This property is seen in all the other experiments, i.e., when p_d is high, the agent avoids the danger zone and exhibits compromise behavior in the lower right region, but when p_d is low, the agent moves straight to the target.

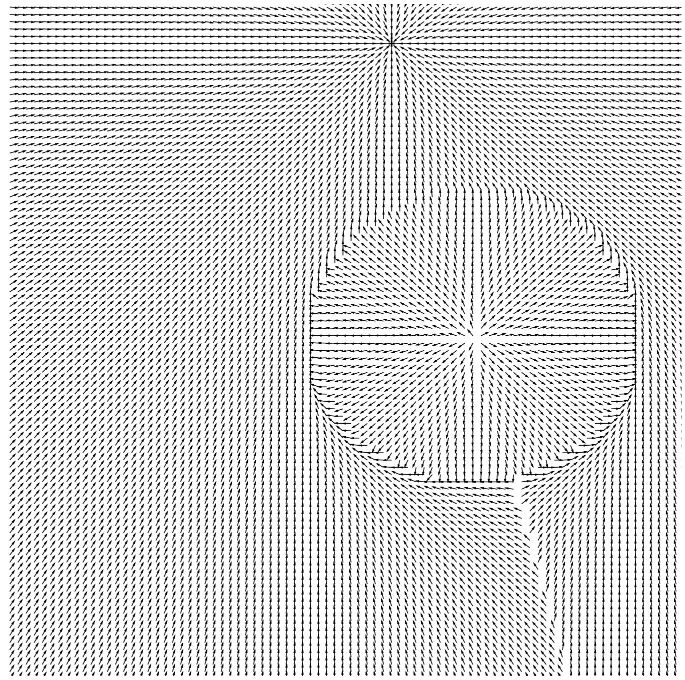

Figure 2: Optimal behavior when the target and danger are likely to stick around ($p_t = 0.995, p_d = 0.99$, and $p_n(d)$ is Linear A).

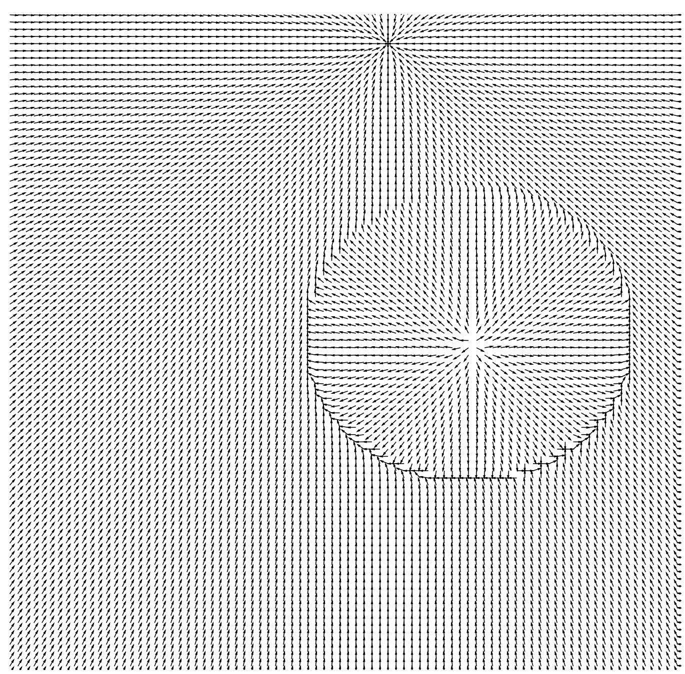

Figure 3: The same scenario as figure 2, but with $p_d = 0.5$. It shows effect of p_d on behavior outside the danger zone.

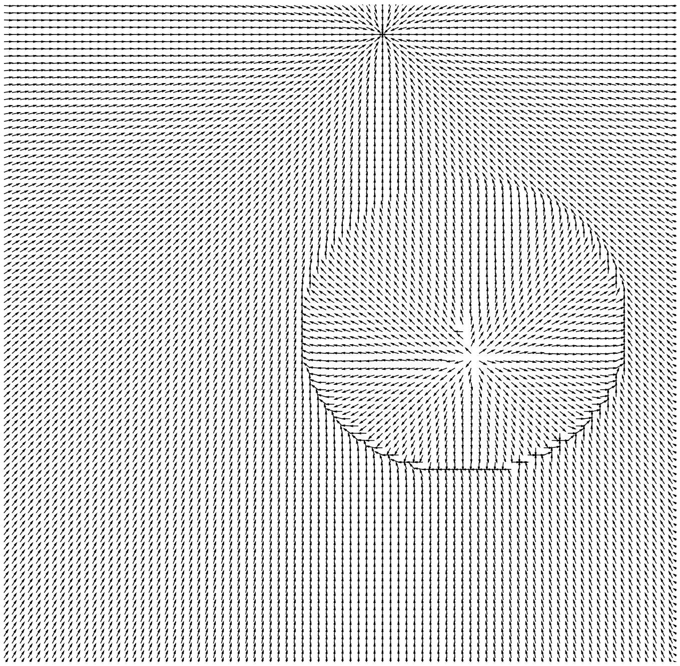

Figure 4: When $p_t = 0.95, p_d = 0.5$, and $p_n(d)$ is Linear A, the results show some compromise action within the danger zone as well as without.

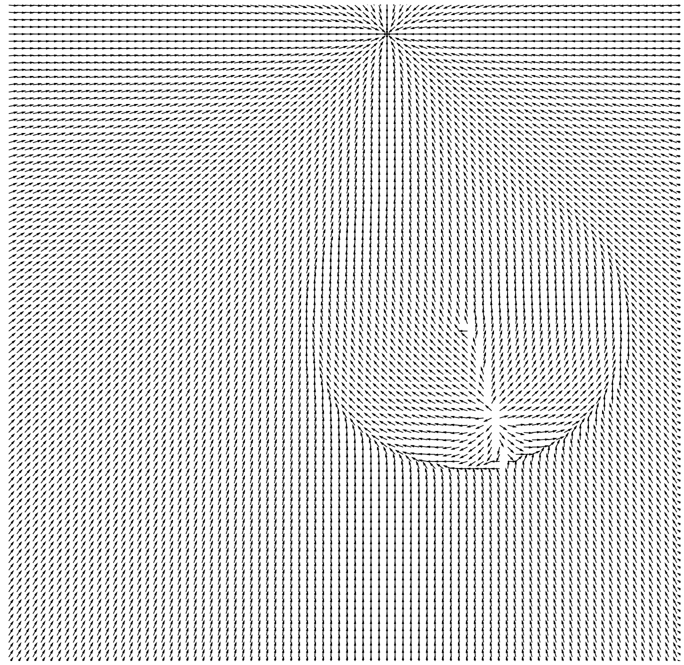

Figure 5: $p_t = 0.95, p_d = 0.5$, and $p_n(d)$ is Linear B.

When $p_t = 0.95, p_d = 0.99$, and $p_n(d)$ is Linear A, the results are qualitatively identical to figure 2, but when $p_t = 0.95, p_d = 0.5$, and $p_n(d)$ is Linear A, we start to see some serious compromise action (figure 4). The combination of both the urgency to get to the target with the likelihood that the danger will disappear leads to more target focused behavior in the danger zone.

When using Linear B, the behavior is identical to Linear A when p_d is high. When p_d is low, the low probability of a strike makes the compromise action more pronounced (figure 5).

With the non-linear $p_n(d)$ functions, compromise action is seen clearly in all cases. Figure 6 shows $p_t = 0.995, p_d = 0.99$, and $p_n(d)$ is sigmoid. The compromise behavior is evident both near the center of the danger zone and again near the edges as the probability of a strike drops gradually from the danger. This is the same for the exponential $p_n(d)$.

The quantitative results of the optimal strategy compared to the greedy strategies described above is shown in table 1. The table shows how the various strategies (optimal, max goal, and skirt) compare to each other in term of percentage improvement. The percentages are of the average expected utility for each strategy across all the starting positions and scenarios[2] listed. “All” is across all scenarios and starting positions; “opposite” is across just the starting positions that are opposite from the target (the lower right region); “danger zone” is across the starting positions inside the danger radius; “Linear A” is all positions when the $p_n(d)$ is Linear A; “Linear B”is all positions when the $p_n(d)$ is Linear B; “Exponential” is all positions when the $p_n(d)$ is Exponential”; and “Sigmoid” is all positions when the $p_n(d)$ is Sigmoid”. The direct strategy was predictably poor (less than half as good as the other strategies across all trials, and 1/6 as good inside the danger zone) so we omitted those results from the table. We see that across all samples, the optimal behavior performs

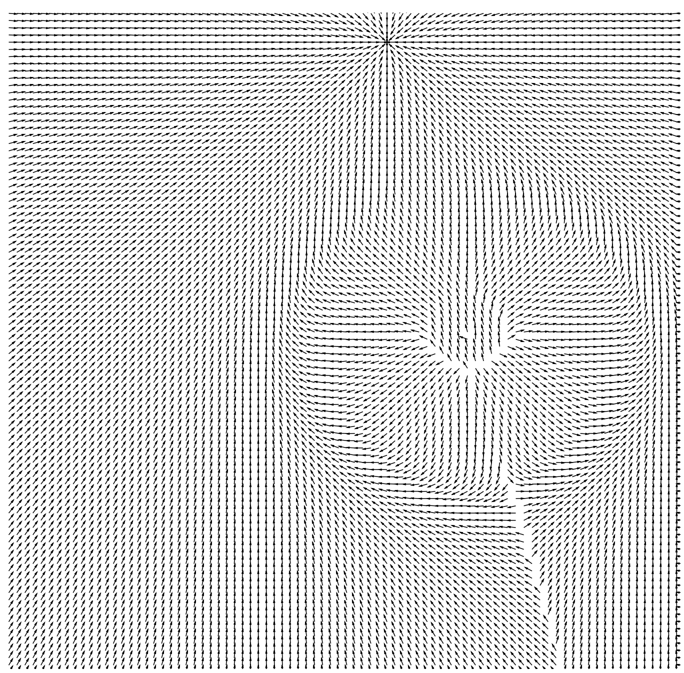

Figure 6: $p_t = 0.995, p_d = 0.99$, and $p_n(d)$ is sigmoid.

[2] A scenario is a single set of values for the parameters in the model.

scenario	optimal over max goal	optimal over skirt	skirt over max goal
all	29.6%	0.1%	29.1%
opposite	64.9%	0.2%	63.3%
danger zone	26.2%	0.01%	26.1%
Linear A	40.9%	0.02%	40.8%
Linear B	13.5%	0.1%	13.1%
Exponential	48.6%	0.03%	48.5%
Sigmoid	16.7%	0.2%	15.2%

Table 1: Results comparing optimal compromise behavior to the greedy strategies.

26% better than max goal, but only 0.1% better than skirt. When we consider just those locations on the other side of the danger zone from the target, we see the benefit is greater for optimal over max goal, but still only slightly so over skirt. This is the same for when we consider just those locations inside the danger zone, or we consider just the samples from each of the $p_n(d)$ functions.

6 Discussion

In discussing the results above, we will present some new insights into the nature of the compromise problem, develop its dual nature, propose a new hypothesis and reinterpret the data from ethology.

6.1 Experimental Results

The biggest surprise in the qualitative results is in the number of scenarios where there is almost no compromise action at all. It appears that in stable environments, the priority is to get away from the danger as soon as possible. Even in a case where the target is likely to disappear and the danger unlikely to remain more than a few time steps, with a moderate chance of a strike, the best thing to do is to flee the danger first (figure 4). In contrast, the probability of a strike has a larger effect on the qualitative behavior than we would have suspected, as shown in figures 5 and 6. The pattern of optimal behavior in figure 6 is as we predicted around the edge of the danger zone, but not at all what we expected in the center, with the optimal behavior ignoring the danger entirely. We are exploring possible causes for this.

The quantitative results in table 1 show that compromise actions in the danger zone (an original reason for proposing them) provide much less benefit than compromise actions in the area opposite from the target. On the other hand, the optimal compromise actions are significantly better than the max goal strategy. This arises from the zig-zag nature of this strategy resulting in much longer paths to the target. When this zig-zag is removed (as in the skirt strategy) the optimal strategy is only the slightest bit better. Although there appear to be other patterns in the data with respect to which locations or which $p_n(d)$ functions favor which strategies, more research need to be done to reach a conclusion.

6.2 Blending vs. Voting

In the work so far, we have looked at compromise candidates using the view described in the quotes and examples from ethology given above. The result is that compromise actions qualitatively appear to be blends of the actions best for each sub-goal (the best direction to move is somewhere in-between the directions that are best for each of the goals). There is an alternative description

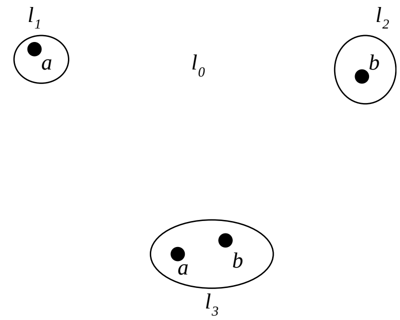

Figure 7: An example scenario where compromise makes sense.

of compromise candidates, also described by Tyrrell, sometimes called the council-of-ministers analogy. In this perspective, there are a collection of "ministers" or experts on achieving each of the agent's various goals. Each minister votes for courses of action that it likes, casting, for example, five votes for its favorite action, four for its second favorite, and so on. The prime minister tallies all the votes and selects the action with the most votes. In this configuration, the compromise selected can be radically different from the non-compromise actions. Imagine an agent at a location l_0 that needs some of resource a and some of resource b. There is a quality source of a at l_1, a location far from a quality source of b at l_2. There is a single low-quality source of both a and b at l_3 (figure 7). Assuming that the utility of a at l_n is a_n, and there is some cost of movement c (a chance of the resource moving away or a direct cost such as energy consumed) then the agent should move l_3 whenever $a_3+b_3-c(\overline{l_0,l_3}) > a_1+b_2-c(\overline{l_0,l_1}+\overline{l_1,l_2})$. In the council-of-ministers, the a minister would cast some votes for l_1, but also some for l_3. Similarly, the b minister would cast votes votes for both l_2 and l_3. The agent might then select moving l_3 as its compromise choice when it is beneficial.

This presents us with an interesting discrepancy: in one model of compromise selection, drawn from real world examples in ethology, compromise is a form of action blending that appears to have little overall benefit to the agent in the prescriptive goal case, and benefit in a limited sense in the proscriptive goal case. In the other, largely hypothetical, model, compromise seems much more granular, results in actions that are qualitatively different from the non-optimal actions, and appears to have the ability to confer real advantage. In the literature, this contrast (in terms of compromise) is unknown, beginning with Tyrrell who used the two definitions interchangeably.

It is our position that the difference between these two models of compromise are because of the level at which the action is defined. Blending compromises take place at the lower levels, where the outputs are essentially the motor commands for the agent. Thus changes allow for little variation in the output. Voting compromises take place at a higher level, where each choice can result in many varied low-level actions. For purposes of distinction, we will call low-level actions[3] *actions* and higher-level actions[4] *behaviors*[5]. Thus selecting a different behavior module

[3]such as *move 1 unit at 2.1 radians*

[4]such as *go to location* l_3

[5]We rely on the behavior-based robotics notion of "behavior" as a reactive module designed to achieve a particular goal. They are also commonly referred to as *goals* or *tasks*. Their important property is higher level of abstraction over actions.

can have wildly different effects at the action level. We believe this distinction was not made by the early researchers in action selection because their experimental environments were entirely discrete and grid-based, thus affording few action options to the agent. For Tyrrell, there was little difference between compromise actions and compromise behaviors.

We note that the "three-layer architectures" in robotics do explicitly make this distinction, where higher layers select between multiple possible behaviors, and then at lower layers, multiple active behaviors select actions [Gat, 1991; Bonasso *et al.*, 1997]. When and where compromise behavior is included varies from instance to instance in an ad hoc manner. Many modern hierarchical action selection mechanisms that explicitly use voting-base compromise tend to do so at the behavior level only [Pirjanian *et al.*, 1998; Pirjanian, 2000; Bryson, 2000].

6.3 The Compromise Behavior Hypothesis

The experiments here and in previous work, with the insights discussed above, lead us to propose the following Compromise Behavior Hypothesis:

> Compromise at the action level confers less overall benefit to an agent than does compromise at the behavior level. Compromise behavior is progressively more useful as one moves upward in the level of abstraction at which the decision is made, for the following reasons:
>
> 1. In simple environments (e.g. two prescriptive goals), optimal compromise actions are similar to the possible non-optimal compromise actions as well as the possible non-compromise actions. As such, they offer limited benefit. In these environments there is no possibility of compromise at the behavior level.
> 2. In complex environments (e.g. where multiple resources are to be consumed in succession such as the scenario depicted in figure 7) compromise behavior can be very different from the active non-compromise behaviors, endowing it with the potential to be greatly superior to the non-compromise.
> 3. In complex environments, optimal or even very good non-optimal actions are prohibitively difficult to calculate.

In the complex environments, optimal compromise *actions* may offer little benefit over actions derived from compromise *behaviors* for the same reason as in 1 above: the optimal action is too similar to the non-optimal action. For example, in the figure 7 scenario, a behavior that decides to move the agent to l_3 can ignore the locations of a and b at l_3 and just generate an action to move to l_3 in general. This non-optimal behavior-generated action will be nearly as good as the optimal action generated by considering the location and qualities of all the a and b, yet the optimal action will come at an enormous computational cost. We propose to begin testing this hypothesis with just this scenario. We predict that the optimal action will be to move toward a location between a and b in l_3, but this optimal action will be essentially just as good as a movement to any other part of l_3.

6.4 Ethological Data Reinterpreted

If it is true that compromise actions are less helpful than compromise behavior, why are so many examples drawn from ethology used to demonstrate compromise actions in animals? It may be that the interpretation of the animal data has been overzealous. In each case there are other possibilities to explain the behavior that do not involve the weighing of compromise actions, or even involve the animal's action selection mechanism at all.

For instance, in the two-prescriptive-goal examples with frogs and crickets following a curved path between two prescriptive goals, an alternative explanation might be that the multiple targets are being merged at the perceptual level, with the ear or auditory system averaging the position of the two targets before any action selection mechanism has an opportunity to consider its options. In this interpretation the behavior would be an accident of morphology, not an attempt to maximize the creature's utility.

Some examples of potential compromise behavior, such as dogs combining a display of fear with one of anger [Lorenz, 1981], or the goose trembling when torn between a prescriptive and a proscriptive goal, may be less an example of compromise behavior, and more a superposition of the two behaviors. This effect arises from the behaviors not sharing a common final path, enabling them to both be expressed simultaneously. In the case of the geese, the resulting behavior is probably one of the least beneficial actions that could be selected, rather than approximating optimality.

Admittedly, these re-interpretations are speculative, but they may not be much more speculative than the idea that they are a result of deliberate consideration of compromise candidates.

7 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 ethologists and AI researchers in the past. In it we have shown that optimal compromise actions in a proscriptive goal case are qualitatively different from what was predicted. They further afford little benefit when compared to a minimally compromise-enabled strategy. We proposed that compromise is not especially useful at the action level, but is useful at the higher behavior level. Future work will revolve around testing, validation or refutation of this Compromise Behavior Hypothesis.

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 N0001404WR20377.

References

[Arkin, 1998] R. Arkin. *Behavior-Based Robotics*. MIT Press, Cambridge, MA, 1998.

[Avila-Garcia and Canamero, 2004] O. Avila-Garcia and L. Canamero. Using hormonal feedback to modulate action selection in a competitive scenario. In *From Animals to Animats 8: Proceedings of the Eighth International Conference on Simulation of Adaptive Behavior (SAB)*, pages 243–254, 2004.

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

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

[Bonasso *et al.*, 1997] R.P. Bonasso, R.J. Firby, E. Gat, D. Kortenkamp, and D. Miller. Experiences with an architecture for intelligent, reactive agents. *Journal of Experimental and Theoretical Artificial Intelligence*, 9(2):237–256, 1997.

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

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

[Caldwell, 1986] G. Caldwell. Predation as a selective force on foraging herons: effects of plumage color and flocking. *Auk*, 103:494–505, 1986.

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

[Choset *et al.*, 2005] H. Choset, K. M. Lynch, S. Hutchinson, G. Kantor, W. Burgard, L. E. Kavraki, and S. Thrun. *Principles of Robot Motion: Theory, Algorithms, and Implementations*. MIT Press, Cambridge, MA, 2005.

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

[Crabbe, 2004] F.L. Crabbe. Optimal and non-optimal compromise strategies in action selection. In *Proceedings of the Eighth International Conference on Simulation of Adaptive Behavior*, 2004.

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

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

[Fraser and Cerri, 1982] D.F. Fraser and R.D. Cerri. Experimental evaluation of predator-prey relationships in a patchy environment: consequences for habitat use patterns in minnows. *Ecology*, 63, 1982.

[Gat, 1991] E. Gat. *Reliable Goal-directed Reactive Control for Real-world Autonomous Mobile Robots*. PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1991.

[Girard *et al.*, 2002] B. Girard, V. Cuzin, A. Guillot, K. N. Gurney, and T. J. Prescott. 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, 2002.

[Grubb and Greenwald, 1982] T.C. Grubb and L. Greenwald. Sparrows and a brushpile: foraging responses to different combinations of predation risk and energy cost. *Animal Behavior*, 30, 1982.

[Howard, 1977] R.A. Howard. Risk preference. In *Readings in Decision Analysis*. SRI International, Menlo Park, CA, 1977.

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

[Hutchinson, 1999] J.M.C. Hutchinson. 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, 1999.

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

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

[Lorenz, 1981] K. Z. Lorenz. *The Foundations of Ethology*. Springer-Verlag, New York, 1981.

[Maes, 1990] P. Maes. How to do the right thing. *Connection Science Journal, Special Issue on Hybrid Systems*, 1, 1990.

[Milinksi, 1986] M. Milinksi. Constraints places by predators on feeding behavior. In T.J. Pitcher, editor, *The behavior of teleost fishes*, pages 236–256. Croom Helm, London, 1986.

[Montes-Gonzales *et al.*, 2000] F. Montes-Gonzales, T. J. Prescott, K. Gurney, M. Humphrys, and P Redgrave. 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, 2000.

[Morris *et al.*, 1978] G. K. Morris, G. E. Kerr, and J. H. Fullard. Phonotactic preferences of female meadow katydids. *Canadian Journal of Zoololgy*, 1978.

[Pirjanian *et al.*, 1998] P Pirjanian, H.I. Christensen, and J.A. Fayman. Application of voting to fusion of purposive modules: An experimental investigation. *Journal of Robotics and Autonomous Systems*, 1998.

[Pirjanian, 2000] P. Pirjanian. Multiple objective behavior-based control. *Journal of Robotics and Autonomous Systems*, 2000.

[Römer, 1993] H. Römer. 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, 1993.

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

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

[Werner, 1994] G. M. Werner. 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, 1994.