Compromise Strategies for Action Selection

Frederick L. Crabbe

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

Abstract

Among many properties suggested for action selection mechanisms, a 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 briefly reviews the history of compromise behaviour and performs experimental analyses of it in an attempt to determine how much compromise behaviour aids an agent. It concludes that optimal compromise behaviour has a surprisingly small benefit over non-compromise behaviour in the experiments performed, it presents some reasons why this may be true and hypothesises cases where compromise behaviour is truly useful. In particular, it hypothesises that a crucial factor is the level at which an action is taken (low level actions are specific, such as "move left leg"; high level actions are vague, such as "forage for food"). The paper hypothesises that compromise behaviour is more beneficial for high-level actions than low-level actions.

Keywords: Action selection; Compromise behaviour; High-level actions; Low-level actions.

1 INTRODUCTION

Agents act. An agent, be it a robot, animal, or piece of software, must repeatedly select actions from a set of candidates. A *controller* is the mechanism within an agent that selects the action. The question of how to design controllers for such agents is the *action selection problem*. Researchers who consider the action selection problem have identified potential properties of these controllers. One such property is the ability to exhibit *compromise behaviour*. A controller exhibits compromise behaviour when the agent has multiple conflicting goals, yet the action selected is not the optimal action for achieving any single one of the those goals, but is good for achieving several of those goals in conjunction. For example, a predator stalking two prey might not move directly toward one of the prey, but in between the two, in case one flees (Hutchinson, 1999). The action would not be optimal for individual goals to catch either prey, yet comes to a compromise between them.

The ability to select a compromise action conveys a benefit to an agent– the optimal action changes in light of other of the agent's goals. There has been disagreement on the beneficiality of compromise behaviour in both animals and artificial agents. This paper investigates the history of compromise behaviour, its various definitions, and the degree to which it (under the most common definitions) confers a behavioural advantage, concluding that the disagreement about the utility of compromise behaviour arises from a fundamentally imprecise notion of what it is. Finally, this paper proposes a new hypothesis for when compromise behaviour is truly beneficial: when the agent selects actions at a high level rather than a low one.

2 BACKGROUND

In order to understand compromise behaviour, it is instructive to examine its history in terms of ethology, comparative psychology, behavioural ecology, artificial intelligence, planning, and robotics. After describing some basic background, this section will describe each of these disciplines in turn, with an emphasis on their relation to the question of compromise behaviour.

2.1 DEFINITIONS OF COMPROMISE

In many approaches to animal behaviour, the full ramifications of an action are weighed in light of the current situation (see section 2.4). Because this can be computationally expensive (see section 2.5) a computational simplification is to divide the action selection problem into subgoals, solve those optimally, and combine the solutions (see section 2.2). It is with respect to this latter strategy that compromise behaviour (acting such that no single subgoal is optimally satisfied) is most often considered (Tyrrell, 1993).

Definitions of compromise behaviour can be categorised on two major dimensions, the level of the action and whether the goals are prescriptive or proscriptive. Each of these dimensions is defined in detail below.

One of the primary characteristics of the different versions of compromise depends upon the abstraction level of the actions selected by the agent. For instance, a low-level action might be for an agent to contract left quadriceps 3 cm. A higher level action might be to transfer itself to a particular location. At the highest levels, an action might be to forage for food, or mate. The distinction is based on the level of specificity given by the action; the first is as specific as possible, while the third leaves flexibility as to how it is to be accomplished. The nature of a particular action selection situation varies based on the level of the actions involved. As will be seen, different authors consider compromise behaviour at different levels.

The other dimension of distinction is the prescriptive or proscriptive nature of the agent's goals. Prescriptive goals are those that are satisfied by the execution of an act, such as the consumption of a resource. Proscriptive goals encourage an agent to not perform certain actions in certain situations. These goals are 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. These goals include avoidance goals, such as remaining at a safe distance from a predator. This paper will not explicitly consider evolutionary goals that are always active for the life of the agent, such as maximising the chance of survival or maximising the chance of reproductive success.

The remainder of this section will review the history of the concept of compromise behaviour from the point of view of the above mentioned fields, demonstrating how the perspectives

of compromise behaviour developed.

2.2 ETHOLOGY

Ethology (the study of animal behaviour) and the study of artificial agents are both concerned with the nature of behaviour and the selection of action. The former considers animal behaviour descriptively and analytically (Tinbergen, 1950), while the latter considers it synthetically via the construction of agents (Pfeifer and Scheier, 1999; Todd, 1992).

Traditionally, the ethologist studies animals in their natural environment, focusing on how they behave in the presence of multiple simultaneous drives. One of the main results of ethology is the identification of fixed action patterns (FAPs) (Brigant, 2005; Dewsbury, 1978; Lorenz, 1981), where an animal exhibits fixed behaviour when it receives a particular type of stimulus. One common example is of the greylag goose (*Anser anser*), which will exhibit a behaviour of rolling an egg back into its nest using a fixed motion pattern, completing the motion pattern even if the egg is removed (Lorenz, 1981). Careful observation of this and similar behaviours led researchers to hypothesise that these individual action patterns are controlled by separate innate modules that compete for expression in the animal's behaviour (Burkhardt, 2004).

The idea that the modules might compete for expression in behaviour led to investigations into how these conflicts might be resolved. Hinde (1966) lists nine different resolution mechanisms observed in animals. These mechanisms include exhibiting just one behaviour, alternating between the multiple behaviours, and compromise behaviour.

Natural compromise behaviour takes multiple forms. First, it is either unimodal or bimodal in the input. In unimodal input, signals from different sensors of the same type (each ear, for example) cause the animal to consider each a separate goal (Lund et al., 1997). Bimodal input combines signal from two different types sensors (eyes and ears, for example) for compromise. In both cases, compromise is typically considered a result of competition for effector mechanisms at a low level (Hinde, 1966). Thus if one FAP controls only leg motion while another only head movement, their simultaneous expression would not be considered compromise behaviour but rather superposition.

Low-level, prescriptive, unimodal compromise has been observed in the crustacean *Armadillium* when performing tropotaxis toward two light sources (Müller, 1925), and katydids when performing phonotaxis (Bailey et al., 1990; Latimer and Sippel, 1987; Morris et al., 1978). The fish *Crenilabris* displays low-level prescriptive, bimodal compromise in its orientation behaviour between its reaction to light and its reaction to the direction of gravity (von Holst, 1935).

Evidence for high-level compromise behaviour in nature is less clear, though it may be argued that it can be seen in blue herons, which select sub-optimal feeding patches to avoid predation by hawks in years when the hawk attacks are frequent (Caldwell, 1986). Similar behaviour has been shown in minnows (Fraser and Cerri, 1982), sparrows (Grubb and Greenwald, 1982), pike and sticklebacks (Milinksi, 1986). Indeed, a great many studies suggest that animals balance the risk of predation against foraging or other benefits (Brown and Kotler, 2004; Lima, 1998).

2.3 COMPARATIVE PSYCHOLOGY

Concurrent with the developments in ethology was a competing branch of study, comparative psychology, that examined many of the same issues (Dewsbury, 1978). This approach differed from ethology in that individual phenomena were studied in isolation, and there was much greater emphasis placed on learning over that of innate mechanisms (Thorpe, 1979). Researchers went to great length to ensure in their experiments that only one drive was active in the test animal (Dewsbury, 1992). This enabled the experimenter to delve deeply into questions about that particular behaviour without interference from others, but it limited investigation into interaction of behaviours. In recent years, the branches of ethology and comparative psychology have been synthesised (Dewsbury, 1992), but early theoretical work had important influences in artifical intelligence (see section 2.5).

2.4 OPTIMAL BIOLOGICAL APPROACHES

Modern trends in biology have employed formal models and optimisation techniques borrowed from decision theory and operations research (Clemen, 1996; Hillier and Lieberman, 2002) in order to determine optimal behaviour. Behavioural ecology is the study of the interaction between an organism's environment and its behaviour, as shaped via natural selection (Krebs and Davies, 1997). Under the assumption that selection optimises behaviour to maximise reproductive success, to understand animal behaviour it is important to analyse it with techniques that optimise objective functions that describe reproductive success. For example, in the field of foraging theory (Stephens and Krebs, 1986), techniques such as linear programming (Hillier and Lieberman, 2002) or dynamic programming (Bertsekas, 2005) are used to find optimal foraging behaviour in terms of such features as maximising energy intake and minimising exposure to predators (Brown and Kotler, 2004; Houston and McNamara, this volume; Lima, 1998; McNamara and Houston, 1994; Seth, this volume). In these studies, optimisation is used as a basis of comparison and explanation toward natural selection; it is not posited as the decision making process the animal itself uses. Optimisation is computationally expensive, such that the time to compute solutions grows exponentially with the complexity. Complicated problems cannot be solved in short periods of time (Bertsekas, 2005).

When examining behavioural choice with these optimal techniques, compromise behaviour is not an explicit issue because the techniques combine the subgoals into a single objective function to be optimised. As such, optimal solutions to the individual sub-goals are not considered, only the solution to the overall objective function.

2.5 ARTIFICIAL INTELLIGENCE AND PLANNING

The field of planning within artificial intelligence was delayed in development until the advent of robotic hardware sufficiently sophisticated to exhibit agent-like behaviour (Fikes and Nilsson, 1971). The approaches used came from the operations research and computer science communities, where the agent attempted to formulate a mathematical proof of the correct action to take in the agent's current situation, with influence from comparative psychology (Newell and Simon, 1976).

A typical planning problem is represented as a conjunction of logical relatively high-level predicates. For example, a hypothetical hospital robot might have a planning goal:

Have(robot,medicine003) $\wedge$ *In(robot,room342)*, indicating that the robot should both be in possession of the medicine and be in the correct hospital room.

From the planning perspective, this is a single goal that is to be achieved by achieving each of its component parts such that there comes a time when both are simultaneously true. The individual predicates, also known as sub-goals, can conflict with each other. For example the robot can take actions so as to make the *In(robot,room342)* to be true while *Have(robot,medicine003)* is false. The robot must then take actions to make *Have(robot,medicine003)* that may in turn

make *In(robot,room342)* false. This conflict is different from the conflict between FAPs in that it arises from the order actions are performed, not in which sub-goal will be achieved. If multiple sub-goals are inherently in conflict such that they cannot both be simultaneously true, then the overall goal is unattainable. Further, because subgoals cannot be partially satisfied (they are simply true or false) it is impossible for the agent to trade in some quality in the satisfaction of one subgoal in order to improve the quality of others. A survey of the state of the art in planning can be found in Gallab et al. (2004).

Other features of the planning problem also bear resemblance to features of the compromise behaviour question. For instance, often a single action can move the agent closer to the satisfaction of more than one of the goal literals. This "positive interference" (Russell and Norvig, 2003) is unlike compromise behaviour however, in that there is nothing lost in the selection of this action. Negated literals in a goal (e.g. $\neg$*In(robot,room342)*) are unlike proscriptive goals in that they they must only be not true at some point for the goal to be satisfied, as opposed to never becoming true.

For the reasons described above, the notion of compromise behaviour was unfamiliar to AI researchers until the 1980s (Brooks, 1986), and optimality under compromise was unexamined.

A major drawback of the AI approaches is that attempts to prove correct actions can be prohibitively expensive in moderately complex environments. If an agent is limited to just ten actions at any time, then each step into the future increases the number of possible outcomes to consider by a factor of ten. If the solution to the current problem is twenty steps long, Then the program must examine on the order of 10^{20} possible sequences (Russell and Norvig, 2003). For comparison, there have been (estimated) 10^{18} seconds since the beginning of the universe (Bridle et al., 2003). This high computation cost prevents using these techniques for planning behaviour with low-level actions, where solutions to problems might be many hundreds of steps long.

2.6 BEHAVIOUR BASED ROBOTICS

Eventually, the inability of robotic systems to solve certain problems of the real world (such as those with multiple simultaneous goals) forced roboticists to re-evaluate their approach. In the real world, agents have conflicting goals that must be selected from, and they must be able to adjust quickly to unforseen events. For instance, the agent may find a previously unknown obstacle or discover that an action did not have the desired effect.

The result of the re-evaluation, *behaviour-based robotics*, borrows from ethology the idea that there are multiple innate behaviours that are triggered by sensory input. In the extreme formulation, advocates maintain that all intelligent behaviour can be constructed out of suites of these competing mechanisms (Arkin, 1998; Brooks, 1986; Brooks, 1997). Some have attempted to explain human-level cognition using similar modular approaches (Carruthers, 2004). One advantage to the behaviour-based approach is that the innate reactive systems do not need to plan with low-level actions, and thus are practical to implement. Another advantage is that conflicting goals can be represented.

Because the approach borrows heavily from the ethological tradition, it has the same concerns. These concerns include how conflicts between innate behaviours can be resolved, and whether compromise behaviour itself an important property for controllers. In 1993, Tyrrell introduced a list of fourteen requirements for action selection mechanisms drawn from ethology. 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) In justifying this rule, Tyrrell used a "council-of-ministers" analogy. In this perspective, there are a collection of "ministers" or experts on achieving each of the agent's goals. Each minister casts votes for courses of action that it predicts will solve the goal with which that minister is associated. For example, it might cast five votes for its highest ranked action, four for the next highest ranked, and so on. The agent then selects the action that receives the most votes. Note that this characterisation of compromise is of high-level compromise. Tyrrell's list has had significant impact on the Action Selection field (Bryson, 2000; Decugis and Ferber, 1998; Girard et al., 2002; Humphrys, 1996), and a number of researchers have developed systems to meet the criteria he set out (Avila-Garcia and Canamero, 2004; Blumberg, 1994; Blumberg et al., 1996; Crabbe and Dyer, 1999; Maes, 1990; Montes-Gonzales et al., 2000; Werner, 1994; Avila-Garcia and Canamero, 2004).

2.7 CURRENT STATUS

Although some researchers in behaviour-based robotics considered it "obviously preferable to combine [the] demand [to avoid a hazard] 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) more recent modelling work generated results that seem to contradict the claim (Bryson, 2000; Crabbe, 2004; Jones et al., 1999), in that artificial agents without the ability to select compromise actions often perform as well on tasks as those that can select compromise actions. If valid, these results suggest that the appropriation of this idea from ethology was not necessary for high performing artificial agents. A central thesis of this paper is that this error occurred in the case of compromise behaviour because it had been poorly defined, in particular, that no distinction was drawn between high-level compromise and low-level compromise. Although low-level compromise is what is seen in much of the action selection literature, its existence was justified by arguments concerning high-level compromise. This equivocation has caused confusion on these topics.

Some artificial agent researchers that use ethological ideas directly are those that design systems not to perform better in the sense of scoring higher on a metric, but to appear more natural to observers. These systems appear in the areas of computer graphics and video gaming, where a naturalistic appearance to a human viewer is necessary to maintain the desired illusion (Iglesias and Luengo, 2005; de Sevin et al., 2001; Thorisson, 1996; Tu, 1996).

Although work mentioned above (Bryson, 2000; Crabbe, 2004; Jones et al., 1999) implies that compromise behaviour is less useful that originally thought, this work is not conclusive. The next section will attempt to analyse the nature of low-level compromise behaviour more thoroughly.

3 EXPERIMENTS

In order to understand the properties of compromise behaviour, it is helpful to examine the optimal behaviour in potential compromise situations. As discussed above, there are multiple formulations of the action selection problem. The experiments here will closely examine those most often described in the ethological and behaviour-based robotics literature, *low-level prescriptive* and *low-level proscriptive*. As the compromise formulations investigated here are low-level, the domain is defined to be that of navigation of a mobile agent, similar to several authors' simulated domains (Maes,

1990; Tyrrell, 1993) or to navigating mobile robots (Choset et al., 2005). In the simulations, space is continuous, but time is discrete, such that the action at each time step is defined as a movement of one distance unit at any angle. Slightly different models are required for each of the proscriptive or prescriptive situations.

3.1 PRESCRIPTIVE EXPERIMENTS

The initial experiments test a scenario where an agent has a goal to be co-located with one of two target locations in the environment. These could be locations of food, water, potential mates, or shelter, etc. At any moment either or both of the targets can disappear from the environment, simulating the intrusion of environmental factors. The agent must select an action that maximises its chances of co-locating with a target before it is removed from the environment.

This scenario is approximated by placing an agent at the origin on a plane. Two targets are placed on the plane, one in the first and one in the second quadrant in the y-range of $(0; 100)$, and x-ranges of $(-100; 0)$ for one target and $(100; 0)$ for the other. Each target will be referred to as t_a and t_b. The agent can sense the location of each target. Sensor information takes the form of complete knowledge of the locations of both targets' (x, y) coordinates. Because the quality of the individual targets may vary, or the types of the targets may be different, the agent has two independent goals to be co-located with them. The strength of the goals are in the range $(0; 100)$. Each goal will be referred to as 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 $1 - p$. Time is divided into discrete, equal sized time-steps. The agent moves at a constant speed, and therefore a constant distance per time step. All distances are measured in the number of time steps it takes the agent to travel that distance. Notationally, $\overline{ij}$ is the distance from some location i to some location j. An agent's action selection problem is to select an angle θ in which direction to move for the next time step. θ is continuous, so the environment is also continuous and the size of the set of actions being selected from is infinite.

Once the agent has executed its action, it is again faced with the same action selection problem. If one of the targets has disappeared, the best action is to move directly to the other target. Compromise behaviour in this task is the selection of any direction to move that is not directly toward either target. Any action selected that is in the direction of one of the targets cannot be a compromise action because it is also the action that is optimal for achieving one of the sub-goals. As the agent repeatedly selects an action, the path it follows resembles a piece-wise linear approximation of a curved path to one of the targets.

3.1.1 FORMAL MODEL

An analysis of compromise candidates is performed 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_i in a state S_j is the sum of the product of the probability of each outcome that could occur and the utility of that outcome:

$$EU(A_i|S_j) = \sum_{S_o \in O} P(S_o|A_i, S_j)U_h(S_o) \tag{1}$$

where O is the set of possible outcome states, $P(S_o|A_i, S_j)$ is the probability of outcome S_o occurring given that the agent takes action A_i in state S_j, and $U_h(S_o)$ is the historical utility of outcome S_o (defined below).

Let $U(t)$ be the utility to the agent of consuming t. Assuming the agent is rational, the set of goals to consume objects will be order isomorphic to the set of the agent's utilities of having consumed the objects. That is, every possible utility corresponds to a matching goal value, such that the order of the utilities from least to greatest is the same as the order of the corresponding goals. Therefore, EU calculated with utilities is order isomorphic with EU calculated with goals instead. For the purposes here, it will be assumed that the goals and utilities are equivalent ($U(t) = G_t$).

A rational agent is expected to select the action with the largest EU. The historical utility of a state is defined as the utility of the state plus future utility, or the maximum of the expected utility of the actions possible in each state:

$$U_h(S) = U(S) + \max_{A_i \in A} EU(A_i|S_j), \tag{2}$$

where A is the set of possible actions. The maximum is found because of the assumption that a rational agent will always act to maximise its expected utility. An agent can calculate EU using multiple actions in the future by recursively applying equations (1) and (2).

Low-level prescriptive compromise behaviour is analysed by comparing an approximation of optimal behaviour with several non-optimal but easy to generate behaviours. The optimal behaviour is approximated based on the dynamic programming technique used by Hutchinson (Hutchinson, 1999). The 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. As with similar dynamic programming techniques, the time to convergence increases as the number and variety of targets increases.

The value calculated is the expected utility of optimal action at an environmental location when the two targets are still remain: $EU(A_\theta|t_a, t_b, \lambda)$. λ is the agent's location in the environment, θ is the angle of the optimal move for the agent and λ' is 1 unit away from λ in direction θ. By equations 1 and 2 the expected utility of being at λ is:

$$\begin{aligned} EU(A_\theta|t_a, t_b, \lambda) =& p^2 EU(A_\theta|t_a, t_b, \lambda') + \\ & p(1-p)EU(A_\theta|t_a, \lambda') + \\ & p(1-p)EU(A_\theta|t_b, \lambda'), \end{aligned} \tag{3}$$

$$EU(A_\theta|t_a, \lambda') = G_a p^{\overline{\lambda' t_a}}, \text{ and,} \tag{4}$$

$$EU(A_\theta|t_b, \lambda') = G_b p^{\overline{\lambda' t_b}}. \tag{5}$$

The total expected utility (equation 3) is the expectation over four possible situations after an action: both targets there, both targets gone, t_a there but t_b gone, and vice versa (the EU of both targets gone is zero). When one of the targets disappears from the environment, the optimal action for the agent to take is to move directly to the other target, as shown in equations 4 and 5. A formal specification of the algorithm is given in the supplementary material.

It is typically computationally prohibitive for an agent to calculate the optimal action using a technique similar to the one described here (the program used for these experiments takes between five and twenty minutes to converge in these two target scenarios). Instead, many researchers propose easy to compute action selection mechanisms that are intended

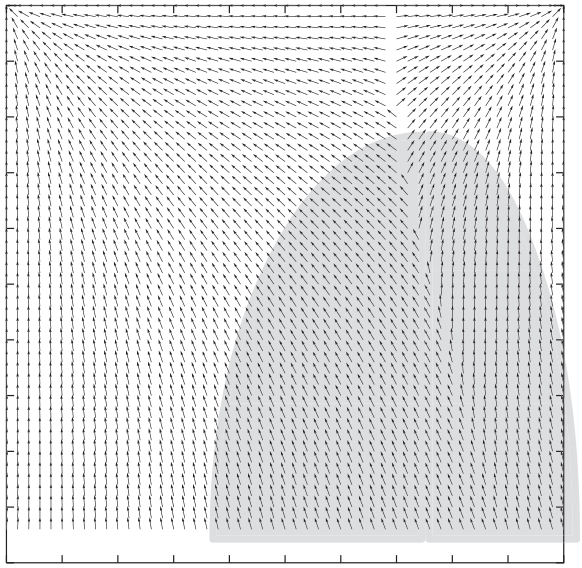


Figure 1: Vector field showing the optimal strategy for a two prescriptive goal scenario. t_a is located in the upper left corner, and t_b is located in the upper right. G_a is 100, G_b is 50. The arrows' lengths are uniform and have no significance. Each arrow direction represents the optimal direction of movement for the agent at that location. A greater proportion of the arrows point toward the upper left, reflecting the larger value of G_a. While the arrows near the targets point directly at the respective targets (indicating non-compromise behaviour is optimal), many of the arrows, especially those in the shaded region in the lower right, point directly at neither target, indicating that in those regions, compromise actions are indeed optimal.

to approximate the optimal action. (Cannings and Orive, 1975; Fraenkel and Gunn, 1961; Houston and McNamara, this volume; Hutchinson and Gigerenzer, 2005; McNamara and Houston, 1980; Römer, 1993; Seth, this volume; Stephens and Krebs, 1986) . 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 behaviour. In the former category, considered here are:

- Closest (C): select the closest target.
- Maximum Utility (MU): select the target with the higher utility.
- Maximum Expected Utility (MEU): select the target with the higher expected utility if it were the only target in the environment (MEU is a non-compromise strategy because it can only select a direction to move that is directly toward one of the targets, and is therefore optimal for one of the agent's sub-goals in isolation).

Of action selection mechanisms that exhibit compromise behaviour, examined here are:

- Forces (F): the agent 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 agent and the target location. Let AngleTo() be a function of two locations that returns the angle from the first location to the second. If V_a is the force vector from T_a, the the direction of V_a is AngleTo(λ, T_a) and the magnitude of V_a is $G_a/\overline{\lambda T_a}^2$. The direction the agent moves (θ) is:

$$\theta = \text{AngleTo}(\lambda, V_a + V_b)$$

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

$$\theta = \text{AngleTo}(\lambda, \arg\max_{\lambda'}(G_a/\overline{\lambda' T_a}^2 + G_b/\overline{\lambda' T_b}^2))$$

- Exponentially Weakening Forces (EWF): This strategy is identical to the forces strategy, except the pulling effects of the targets falls exponentially with distance, rather than quadratically. The magnitudes of the two vectors are $G_a p^{\overline{\lambda T_a}}$ and $G_b p^{\overline{\lambda T_b}}$. It is predicted that since expected utility falls exponentially with distance, this strategy may perform better than forces.

The expected utility of each of these non-optimal 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 (further experiments and results that describe the effects of the initial parameters on compromise can be found in the supplementary material).

3.1.2 PRESCRIPTIVE RESULTS

The results reported here are based on 50,000 scenarios. Each scenario was a set of parameters (G_a, G_b, t_a, t_b) selected randomly from a uniform distribution. The simulations were written in Lisp, compiled in Franz Allegro Common Lisp, version 7.0, and run on a cluster of twenty five Sun Blade 1500s, for 347 computer-days. For each scenario, the expected utility of each of the action selection mechanisms described in the previous section were computed: closest (C), maximum utility (MU), maximum expected utility (MEU), forces (F), signal gradient (SG), and exponentially weakening forces (EWF). The expected utilities of optimal behaviour using the dynamic programming technique were computed (an example of the optimal behaviour is shown in figure 1). Table 1 compares the three non-compromise mechanisms (C, MU, and MEU), using the worst performer (MU) as a baseline. The table reports the average percentage improvement of the strategy over MU (e.g. the closest strategy performs on average 9% better than the maximum utility strategy). It also reports the percentage of cases where the strategy selected the correct action of the two possible. MEU is the best of the three as it selects the better target in most cases and its overall expected utility is 15% better than MU. MEU selects the worse target only 0.68% of the time. The table also shows that Closest is a better strategy than Maximum Utility. This may be so because the expected utility of a target falls exponentially with distance, so that closer targets have higher expected utility than targets with higher raw utilities.

Table 2 compares the compromise based mechanisms with the best non-compromise strategy, MEU. It shows both the average percentage improvement improvement over MEU and the percentage improvement over MEU in the single best scenario. There are three important aspects of this table. The

	MU	C	MEU
% correct	70.29	79.99	99.32
% over MU	0.0	9.35 (3.97)	15.31 (3.44)

Table 1: Comparison of non-compromise strategies. Each strategy is listed in the column. The first row reports the percentage of times that strategy makes the correct selection. The second row reports the strategy's performance improvement over MU. Numbers in parentheses are the standard deviation. The numbers are generated from the 50,000 trials. Percentage improvement is calculated as (Score of Strategy A - Score of MU)/Score of MU . All differences were significant on a Mann-Whitney U-test to a confidence of 0.999.

	F	SG	EWF	Optimal
% over MEU	-4.07	-2.79	-2.47	1.12
% over MEU (best case)	4.84	4.82	20.56	22.73

Table 2: Comparison of compromise strategies to the non-compromise maximum expected utility (MEU). For each strategy, the average expected utility improvement over MEU is given in the first line. The best expected utility improvement over MEU in a single scenario is provided in the second line. The optimal strategy performs the best, but is less beneficial than expected. The non-optimal strategies (F, SG, and EWF) are all worse than the non-compromise MEU, though EWF is the best of that set. Differences in the averages were significant on a Mann-Whitney U-test to a confidence of 0.99.

first is that the optimal strategy is only 1.1% better than the non-compromise based MEU. This contradicts the intuition (discussed above) that optimal behaviour would be significantly better than a non-compromise approach. The result is consistent however, with the non-continuous space experiments of Crabbe (2002) and the study in Hutchinson (1999).

The second important aspect is that all of the non-optimal compromise based strategies performed *worse* than the MEU strategy. These results may help explain why some researchers have found that compromise behaviour is unhelpful (Jones et al., 1999; Bryson, 2000; Crabbe, 2004): the commonly used *tractable* compromise strategies perform worse than a non-compromise strategy.

The final aspect of table 2 to note is that EWF is the best performing of the easy to compute compromise strategies tested. While it is 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. Examining the score for the *best* scenario for EWF shows that it is nearly as high as the best scenario for optimal.

3.1.3 Prescriptive Discussion

With respect to animals and natural action selection, the results presented here imply that animals that exhibit low-level prescriptive compromise behaviour are either: behaving non-optimally; using an as yet unproposed compromise based action selection strategy; or behaving in that manner for reasons other than purely to compromise between two targets. Hutchinson (1999) suggests three possible reasons for what appears to be low-level prescriptive compromise behaviour: 1) a desire to not tip off potential prey that it is being stalked, 2) it is a part of a strategy to gather more sense data before committing to a target, or 3) that computational issues yield simple mechanisms that exhibit compromise style behaviour. Hutchinson's reasons are particularly interesting in light of MEU being the best non-compromise strategy. This strategy requires not only detailed knowledge of the targets' locations and worths, but also that the agent knows p. It may be that apparent low-level compromise is an attempt to gather more information about the targets, or that, lacking knowledge of p, animals are unable to use the MEU strategy, in which case the compromise signal gradient or EWF strategies might be the best (although results from foraging theory suggest that animals are able to estimate p accurately (Stephens and Krebs, 1986)). Per Hutchinson's third suggestion, Houston et al. (this volume) suggest that behavioural characteristics can be "side-effects" of rules that evolved in environments that differ from where they are being used, or that the objective function and criteria being maximised are more complex than the scenarios in which they are being tested.

Ghez, et al. (1997) showed that when humans performed a reaching task, a narrow angle between targets led to low-level compromise behaviour, while a wide angle did not. They hypothesise that for widely separated targets, the brain treated each as a separate concept or category, but that for narrowly separated targets, the brain is unable to tease them apart, thus reacting to their superposition. By analysing reaction time, Favilla (2002) showed that humans do appear to be switching mental strategies when changing between compromise and non-compromise behaviour, even when the tasks remain the same. These results may indicate that low-level compromise is a side-effect of other computational mechanisms.

With respect to higher-level actions, the behavioural ecology evidence is much less clear that natural compromise is occurring. For instance in the cases of an animal sub-optimal feeding patches to avoid heightened predator activity, this behaviour could be explained by the animal downgrading the quality of the feeding patch (the G_x) because of the presence of the predators. The animal then compares the utilities of the two directly rather than considering compromise behaviour (Stephens and Krebs, 1986). Alternatively, the animal could be abstracting the problem so that it might be solved optimally.

3.2 PROSCRIPTIVE EXPERIMENTS

Although in the low-level prescriptive experiments compromise behaviour had less benefit than predicted, it could be argued that the prescriptive case is not best suited for eliciting positive results. It may be that compromise is more useful in cases where 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)

This section tests this claim by performing experiments similar to the prescriptive case, but with one proscriptive goal. In these experiments, the environment contains a target and a danger in fixed locations. The danger can "strike" the agent

from a limited distance. The agent has a prescriptive goal to be co-located with the target, and a proscriptive goal to avoid being struck by the danger.

3.2.1 FORMAL MODEL

The model described in section 3.1 requires modification to match this new scenario. The two environmental objects, the target (t) and the danger (d) are treated separately with individual probabilities of remaining in the environment (p_t and p_d, respectively). At each time step, there is a probability $p_n(\lambda)$ 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, calculated from the agent's position λ. The experiments use four different versions of the $p_n(\lambda)$ function. 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. Other notation remains the same.

The application of equations (1) and (2) calculate the expected utility of being at λ:

$$\begin{aligned} EU(A_\theta|t,d,\lambda) =& p_t p_d p_n(\lambda) EU(A_\theta|t,d,\lambda') + \\ & p_t(1-p_d) EU(A_\theta|t,\lambda') + \\ & p_d(1-p_n(\lambda))G_d + \\ & (1-p_t)p_d p_n(\lambda) EU(A_\theta|d,\lambda') \end{aligned} \tag{6}$$

$$EU(A_\theta|t,\lambda) = G_t p^{\overline{\lambda t}}, \text{ and,} \tag{7}$$

$$\begin{aligned} EU(A_\theta|d,\lambda) =& p_n(\lambda') p_d EU(A_\theta|d,\lambda') + \\ & (1-p_n(\lambda'))G_d. \end{aligned} \tag{8}$$

The total expected utility (equation 6) 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 (7). When the target disappears but the danger remains, the agent must flee to a safe distance from the danger, as in equation (8). 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. The area inside the danger radius is the danger zone.

In addition to the optimal strategy described above, three other action selection strategies are examined:

- MEU: The agent moves in accordance with the maximum expected utility strategy, as described in section 3.1.1. Movement is directly to the target, ignoring the danger, because the target has the higher utility. This is a non-compromise strategy that could be expected to do poorly.

- Active goal: This strategy considers only one goal at a time: the danger when in the danger zone and the target otherwise. Using this, the agent moves directly to the target unless within the danger zone. Within the danger zone, the agent moves directly away from the danger until it leaves the zone. This strategy zig-zags along the edge of the danger zone as the agent moves toward the target. Active goal is also a non-compromise 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 such a position, the agent moves along the tangent edge of the danger zone until it can resume heading directly to the target. Skirt is primarily a non-compromise strategy. Outside the danger radius, the agent moves straight to the target. Inside the danger radius the agent moves straight away from the danger.

The expected utility of each of these non-optimal mechanisms can be calculated for any particular scenario by using equations (6), (7) and (8), as in the previous experiments.

For these experiments, four $p_n(\lambda)$ functions were used, all of with with a danger radius of 20:

- Linear A: $p_n(\lambda) = 0.04 \times \overline{d\lambda} + 0.2$ when $\overline{d\lambda} \leq 20$, 1 otherwise.

- Linear B: $p_n(\lambda) = 0.005 \times \overline{d\lambda} + .9$ when $\overline{d\lambda} \leq 20$, 1 otherwise.

- Quadratic: $p_n(\lambda) = (\overline{d\lambda})^2/400$ when $\overline{d\lambda} \leq 20$, 1 otherwise.

- Sigmoid: $p_n(\lambda) = 1/(1 + 1.8^{10-\overline{d\lambda}})$ everywhere.

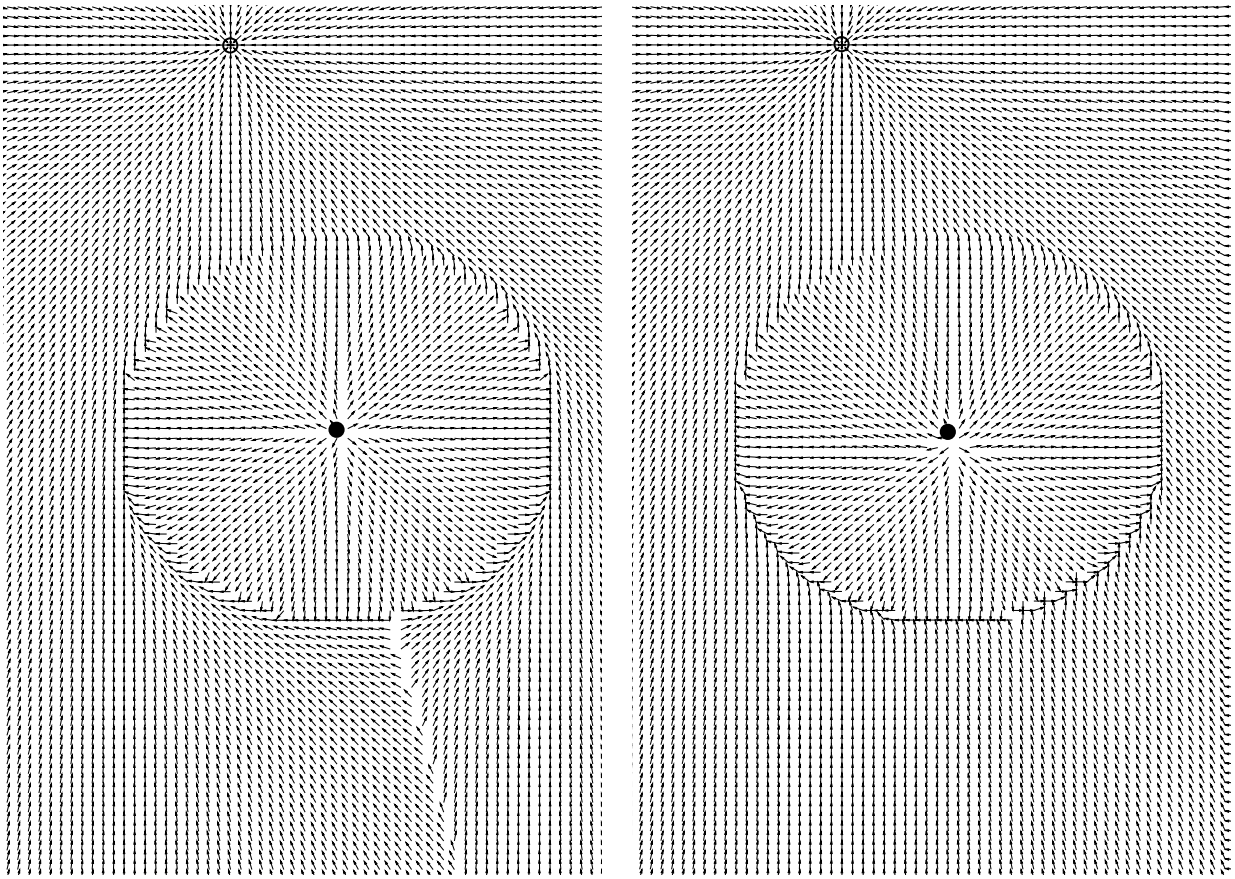
Linear A was a baseline strategy where the probability of a strike was high near the danger, but low at the edge of the danger zone. Linear B makes the chance of a strike low overall, thus increasing the tendency of the agent to remain in the danger zone. This may generate more compromise behaviour. Quadratic has a high probability of a strike for much of the danger zone, but drops off sharply at the edge. This may encourage compromise behaviour near the edge of the danger radius but not at the centre. Sigmoid should resemble quadratic, 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.

1000 scenarios were generated with a target at $(50, 90)$ with a $G_t = 100$ and a danger at $(60, 50)$ with $G_d = -100$. p_t was varied systematically in range $[0.95; 1)$ and p_d was varied systematically in the range $[0.5; 1)$ (These ranges were selected because they contain the most interesting behaviour. For instance, when p_t is too low, the probability that an agent will reach the target quickly approaches zero. Related studies (see (Crabbe, 2002) and supplementary material) indicated that compromise behaviour was greater when $p_t > 0.95$). 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, for 200,000 data points calculated over 312 computer-days.

3.2.2 PROSCRIPTIVE RESULTS

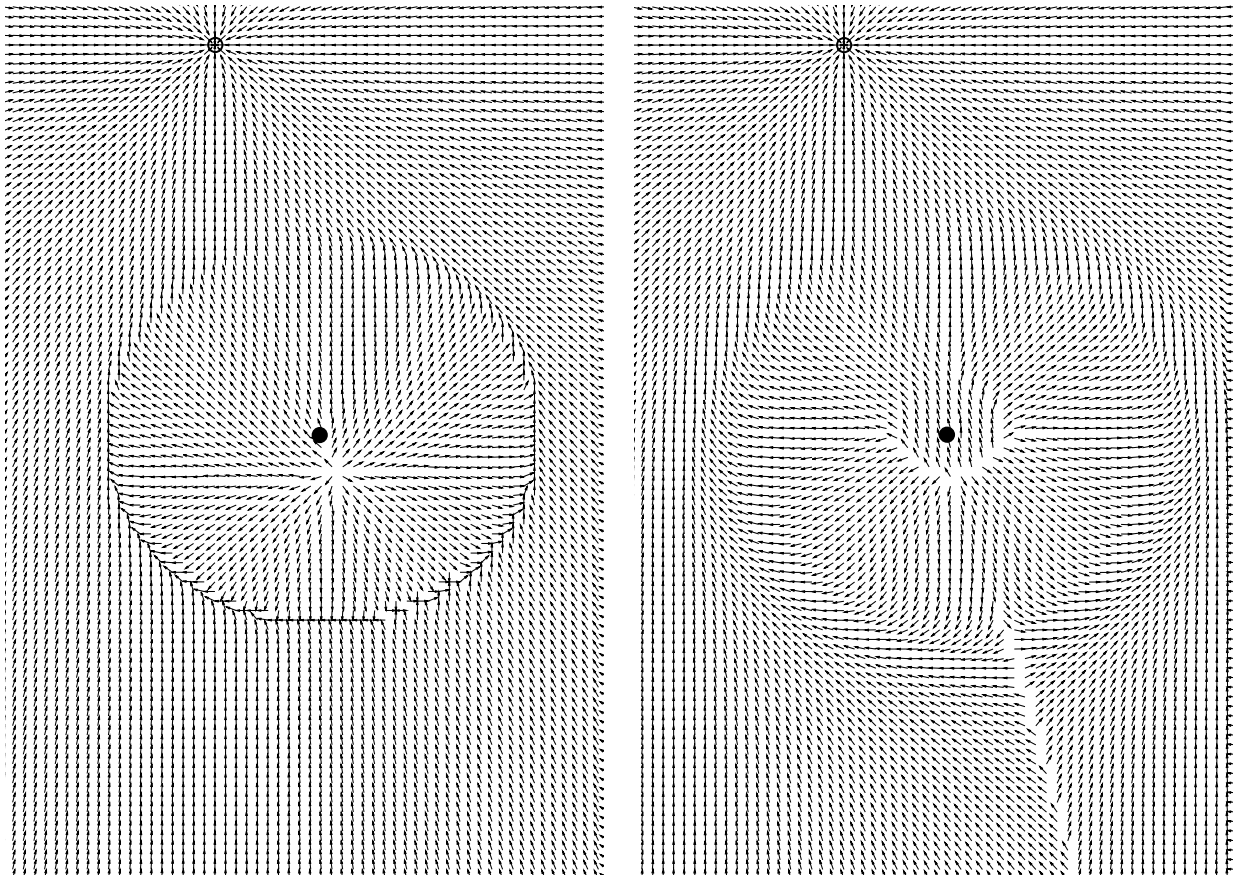
Figure 2(a) shows the results of the optimal strategy when $p_t = 0.995, p_d = 0.99$, and $p_n(\lambda)$ is Linear A. Within the danger zone, there is little display of compromise action; the agent flees directly away from the danger at all locations, ignoring the target. There is compromise action displayed outside the danger zone, to the lower right. The vectors point not at the target, but along the tangent of the danger zone. This phenomenon occurs because the agent moves along the shortest path around the danger zone to maximise the likelihood that the target will remain in the environment until the agent arrives. This compromise in the lower right does not match the common implementations of compromise action. In many architectures, the goal to avoid the danger would not be active when the agent is in that area of the environment (since the agent is too far away from the danger) (Arkin, 1998; Brooks, 1986). Thus one would expect it to have no effect of the action selected.

When p_d is reduced to 0.5 the compromise action in the lower right is less pronounced (figure 2(b)). The optimal



(a) (b)

Figure 2: Optimal behaviour for the agent in proscriptive scenarios. The vectors point in the direction of optimal movement at each given location. As before, the all vectors are unit length. Each target is located along the top of the plot, indicated by an open circle. Each danger is located near the centre of the plot, indicated by a filled circle. The large circle on the plots is formed by the vectors pointing away from the danger in areas inside the danger zone. Outside the danger zone, the vectors point toward the target. In (a) the probability of the target remaining is high ($p_t = 0.995$), the probability of the danger remaining is high ($p_d = 0.99$) and $p_n(\lambda)$ is Linear A. In most locations, the optimal strategy is to move directly toward the target. Inside the danger zone, the optimal strategy is to move directly away from the danger. On the side of the danger zone opposite the target, the optimal strategy is to move to the target along the shortest path while not entering the danger zone. In (b), the parameters are the same as in (a), but the probability of the danger remaining is low ($p_d = 0.5$). The optimal behaviour is qualitatively the same as for (a) except in the area on the side of the danger zone opposite the target, where optimal behaviour is to move directly toward the target (see text).



(a) (b)

Figure 3: Optimal behaviour in situations where more compromise behaviour is shown. The figure should be interpreted in the same manner as with figure 2. In (a) both the probability of the target and the danger remaining are low ($p_t = 0.95, p_d = 0.5$) and $p_n(\lambda)$ is Linear A. Compromise behaviour is evident inside the danger zone because the optimal direction of movement is no longer directly away from the danger. In (b) the probability of the target remaining is high ($p_t = 0.995$), the probability of the danger remaining is high ($p_d = 0.99$), but $p_n(\lambda)$ is Sigmoid. Within the centre of the danger zone, the optimal behaviour is to ignore the danger and move directly toward the target. In areas closer to the edge of the danger zone, the optimal strategy shifts to moving away from the danger. At the extremity of the danger radius, the optimal strategy gradually shifts back to moving toward the target.

strategy is to act as if the danger will disappear before the agent enters the danger zone. This property is seen in all the other experiments, i.e., when p_d is high, optimal behaviour avoids the danger zone and exhibits compromise behaviour in the lower right region, but when p_d is low, the agent moves straight to the target in that region.

When lowering p_t to 0.95 (with $p_d = 0.99$, and $p_n(\lambda)$ as Linear A), the results are qualitatively identical to figure 2(a) (This and other additional plots can be found in the supplementary material). When $p_t = 0.95, p_d = 0.5$, and $p_n(\lambda)$ is Linear A, (i.e. low p_t and low p_d) predicted compromise behaviour emerges (figure 3(a)). The combination of both the urgency to get to the target with the likelihood that the danger will disappear leads to more target focused behaviour in the danger zone.

Examining the non-linear $p_n(\lambda)$ functions, compromise action is seen clearly in all cases. Figure 3(b) shows $p_t = 0.995, p_d = 0.99$, and $p_n(\lambda)$ is sigmoid. The compromise behaviour is evident both near the centre of the danger zone and again near the edges as the probability of a strike drops gradually from the danger. This also occurs when the $p_n(\lambda)$ is quadratic (see the supplementary material).

Comparison between the optimal strategy and the other strategies described above is shown in table 3. The table uses active goal as a baseline and compares skirt and the optimal strategy to it. The MEU strategy was poor (less than half as good as the other strategies across all trials, and 1/6 as good inside the danger zone), so was omitted from the table. The percentages are of the average expected utility for each strategy across all the starting positions and scenarios (200,000 data points). "All" is across all scenarios and starting positions; "opp" is across just the starting positions that are opposite from the target (the lower right region); "DZ" (danger zone) is across the starting positions inside the danger radius; "Lin A" is all positions when the $p_n(\lambda)$ is Linear A; "Lin B"is all positions when the $p_n(\lambda)$ is Linear B; "Quad" is all positions when the $p_n(\lambda)$ is Quadratic; and "Sig" is all positions when the $p_n(\lambda)$ is Sigmoid. Across all samples, the optimal behaviour performs 29.6% better than active goal, but skirt is nearly as good, performing 29.1% better than active goal. When considering just those locations on the other side of the danger zone from the target, the benefit is greater for optimal over active goal, but still only slightly so over skirt. This trend continues for locations inside the danger zone, and samples from each of the $p_n(\lambda)$ functions.

scenarios	optimal over active goal	skirt over active goal
all	29.6%	29.1%
opp	64.9%	63.3%
DZ	26.2%	26.1%
Lin A	40.9%	40.8%
Lin B	13.5%	13.1%
Quad	48.6%	48.5%
Sig	16.7%	15.2%

Table 3: Results comparing optimal and skirt strategies to active goal. The values reflect the average percent expected utility improvement of the the samples. The rows are subsets of the data set. The first row is across all the data points. The second row is just those points in a region on the side of the danger opposite the target. The third row is those points within the danger zone. The fourth through seventh rows are for each of the $p_n(\lambda)$ functions, across all locations in the environment. The data consistently shows that the skirt strategy is nearly as good as the optimal one in all cases. All differences over active goal were significant on a Mann-Whitney U-test to a confidence of 0.999.

3.2.3 PROSCRIPTIVE DISCUSSION

An examination of figures 2 and 3 reveals properties of the optimal strategy that were not initially predicted (see section 3.2.1). In stable environments (figure 2), the priority is to flee the danger. Even in cases 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 optimal action is to flee the danger first (figure 2(b)).

The pattern of optimal behaviour in figure 3(b) is as predicted around the edge of the danger zone (i.e. the Sigmoid $p_d(\lambda)$ function generates low-level compromise behaviour that gradually disappears as one moves farther from the danger), but not at all what was expected in the centre (fleeing the danger directly), with the optimal behaviour ignoring the danger entirely. This occurs because in that region the probability of a strike falls very little as the agent moves away from the danger, yet the probability of reaching the target still falls exponentially with distance. The agent is likely to be struck no matter action it takes, so its best course of action is to move toward the target.

While low-level compromise is shown to be beneficial in the proscriptive experiments, the experiments also show that it is not beneficial in the manner expected, namely that instead of inside the danger zone, low-level compromise is most beneficial outside the danger zone. Indeed, the comparison between the optimal and the skirt strategies shows that the majority of the benefit comes not from finding a compromise between two goals, but from preventing the oscillation between acting on each goal, thus generating longer than necessary paths along the edge of the danger zone. In the cases where the transition at the edge of the danger zone was less behaviourly severe (i.e. the $p_n(\lambda)$ was unlikely to generate a strike, so that optimal behaviour just inside and just outside the zone are similar– when $p_n(\lambda)$ is Linear B or Sigmoid– then the benefit of the optimal strategy is only thirteen to eighteen percent greater than the active goal strategy that zig zags in and out of the danger zone.

4 FINAL DISCUSSION

This paper has presented two sets of experiments analysing low-level compromise behaviour. The experimental setup was based on situations predicted to be amenable toward good compromise actions (Tyrrell, 1993), and using environments that are commonly seen in the artificial agent community (Blumberg et al., 1996). The results show that compromise was not as beneficial as predicted in the prescriptive cases, and while it was beneficial in the proscriptive cases, it: a) took forms different from what was expected, and b) the vast majority of benefit came from low-level compromise that served primarily to shorten the overall path of the agent. This section will discuss the implications of these findings.

4.1 HIGH VS. LOW-LEVEL ACTIONS

Mounting experimental evidence (in this paper and in others (Bryson, 2000; Crabbe, 2004; Jones et al., 1999)) appears to show that compromise behaviour is less helpful than predicted, and yet the intuition that compromise must have greater impact can still be strong. A simple thought exper-

iment makes it appear even more so. 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. Let the utility of a at location l_n be a_n. If 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 to l_3 whenever $a_3 + b_3 - c(\overline{l_0 l_3}) > a_1 + b_2 - c \times \min(\overline{l_0 l_1} + \overline{l_1 l_2}, \overline{l_0 l_2} + \overline{l_1 l_2})$. Using the council-of-ministers analogy, 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.

The key difference between the scenario just described and the experiments described in earlier sections is the nature of the actions selected. The experiments closely resembled the sort of compromise shown often in the ethological literature, where the actions selected appear as a continuous blend of the non-compromise actions, whereas the justification for compromise was posed as a discrete voting system. With voting, the compromise action selected can be radically different from the non-compromise actions.

This difference arises due to the level at which the action is defined. Blending compromises take place at the lower levels, where the outputs are 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. Although this distinction is highlighted here, it is not common in literature. Tyrrell (Tyrrell, 1993) for example used the two definitions interchangeably (it may be that this distinction was not made by the early researchers in action selection in part because their experimental environments were entirely discrete and grid-based, thus affording few action options to the agent). As discussed in section 2, selection of optimal actions at the low level is much more computationally difficult than selecting actions at a high level.

It should be noted that the "three-layer architectures" in robotics do explicitly make this action level distinction, where higher layers select between multiple possible high-level behaviours, and then at lower layers, active behaviours select low-level actions (Bonasso et al., 1997; Gat, 1991). In existing systems, when and where compromise behaviour 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 behaviour level only (Bryson, 2000; Pirjanian, 2000; Pirjanian et al., 1998).

4.2 COMPROMISE BEHAVIOUR HYPOTHESIS

The experiments and insights discussed above, lead us to propose the following Compromise Behaviour Hypothesis:

"Compromise at low levels confers less overall benefit to an agent than does compromise at high levels. Compromise behaviour 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 higher levels. **(2)** In complex environments (e.g. where multiple resources are to be consumed in succession such as the hypothetical scenario depicted in the previous section) good compromise behaviour can be very different from the active non-compromise behaviours, endowing it with the potential to be greatly superior to the non-compromise. **(3)** In complex environments, optimal or even very good non-optimal low-level actions are prohibitively difficult to calculate, whereas good higher level action are not. Furthermore, easy to compute heuristics (such as Forces) are unlikely to generate the radically different actions required for good compromise."

This hypothesis predicts that compromise behaviour will be beneficial in more complex environments, where the computational cost of selecting an action at a low level is prohibitive. In these environments, action selection at a high level, with compromise, may be the best strategy.

5 CONCLUSION

The notion of compromise behaviour has been influential in the action selection community despite disagreements about what precisely it might be. By examining the most common forms of compromise behaviour described by ethologists or implemented by computer scientists (low-level prescriptive and proscriptive), this paper adds credence the idea that while it may exist in nature, low-level compromise behaviour affords little benefit. The paper proposes that compromise is not especially useful at the low levels, but *is* useful at higher levels. Future work will revolve around testing, validation or refutation of this Compromise Behaviour Hypothesis.

Acknowledgements

We would like to thank Chris Brown and Rebecca Hwa for wonderful discussions, and Pauline Hwa for editing advice. We would also like to thank the editors and anonymous reviewers for many helpful comments. This work was sponsored in part by a grant from the Office of Naval Research, number N0001404WR20377.

References

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

Avila-Garcia, O. and Canamero, L. (2004). 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*, pages 243–254.

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 female choice or passive attraction. *Animal Behavior*, **40**:33–42.

Bertsekas, D. P. (2005). *Dynamic Programming and Optimal Control*, volume 1. Athena Scientific, Belmont, MA.

Blumberg, B. M. (1994). Action-selection in hamsterdam: Lessons from ethology. In *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, pages 108–117.

Blumberg, B. M., Todd, P. M., and Maes, P. (1996). No bad dogs: Ethological lessons for learning in hamsterdam. In *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, pages 295–304.

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

Bridle, S. L., Lahav, O., Ostriker, J. P., and Steinhardt, P. J. (2003). Precision cosmology? not just yet. *Science*, **299**:1532.

Brigant, I. (2005). The instinct concept of the early Konrad Lorenz. *Journal of the History of Biology*, **38**(3):571 – 608.

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

Brooks, R. (1997). From earwigs to humans. *Robotics and Autonomous Systems*, **20**(2-4):291–304.

Brown, J. S. and Kotler, B. P. (2004). Hazardous duty pay and the foraging cost of predation. *Ecology Letters*, **7**:999–1014.

Bryson, J. (2000). Hierarchy and sequence vs. full parallelism in action selection. In *From Animals To Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, pages 147–156.

Burkhardt, R. W. (2004). *Patterns of Behavior: Konrad Lorenz, Niko Tinbergen, and the Founding of Ethology*. Chicago University Press, Chicago.

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

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.

Carruthers, P. (2004). Practical reasoning in a modular mind. *Mind and Language*, **19**(3):259–278.

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

Clemen, R. (1996). *Making Hard Decisions: An Introduction to Decision Analysis*. Duxbury Press, Belmont CA.

Crabbe, F. L. (2002). Compromise candidates in positive goal scenarios. In *From Animals to Animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, pages 105–106.

Crabbe, F. L. (2004). Optimal and non-optimal compromise strategies in action selection. In *From Animals to Animats 8: Proceedings of the Eighth International Conference on Simulation of Adaptive Behavior*, pages 233–242.

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*.

de Sevin, E., Kallmann, M., and Thalmann, D. (2001). Towards real time virtual human life simulations. In *Computer Graphics International*, pages 31–34.

Decugis, V. and Ferber, J. (1998). An extension of Maes' action selection mechanism for animats. In *From Animals to Animats 5: Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, pages 153–158.

Dewsbury, D. A. (1978). What is (was?) the "fixed action pattern?". *Animal Behaviour*, **26**:310-311.

Dewsbury, D. A. (1992). Comparative psychology and ethology: A reassessment. *American Psychologist*, **47**:208–215.

Favilla, M. (2002). Reaching movements: mode of motor programming influences programming by time itself. *Exp Brain Res*, **144**:414–418.

Fikes, R. E. and Nilsson, N. J. (1971). STRIPS: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, **2**(3-4):189–208.

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

Fraser, D. and Cerri, R. (1982). Experimental evaluation of predator-prey relationships in a patchy environment: consequences for habitat use patterns in minnows. *Ecology*, **63**(2):307–313.

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

Ghallab, M., Nau, D., and Traverso, P. (2004). *Automated Planning: Theory & Practice*. Morgan Kaufmann Publishers Inc., San Francisco, CA.

Ghez, C., Favilla, M., Ghilardi, M. F., Gordon, J., Bermejo, R., and Pullman, S. (1997). Discrete and continuous planning of hand movements and isometric force trajectories. *Exp Brain Res*, **115**:217–233.

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 *From Animals to Animats 7: Proceedings of the seventh international conference on simulation of adaptive behavior*, pages 75–84.

Grubb, T. and Greenwald, L. (1982). Sparrows and a brush-pile: foraging responses to different combinations of predation risk and energy cost. *Animal Behavior*, **30**:637–640.

Hillier, F. S. and Lieberman, G. J. (2002). *Introduction to Operations Research*. McGraw-Hill, New York.

Hinde, R. A. (1966). *Animal Behaviour*. McGraw-Hill, New York.

Houston, A. and McNamara, J. (2006). Do we expect natural selection to produce rational behavior. *Phil. Trans. Roy. Soc.*

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, Cambridge, UK.

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.

Hutchinson, J. M. C. and Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural Processes*, **69**:97–124.

Iglesias, A. and Luengo, F. (2005). New goal selection scheme for behavioral animation of intelligent virtual agents. *IEICE Trans. Inf. and Syst.*, **E88-D**(5):865–871.

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):27–42.

Krebs, J. and Davies, N., editors (1997). *Behavioural Ecology: an evolutionary approach*. Blackwell Publishers, Oxford.

Latimer, W. and Sippel, M. (1987). Acoustic cues for female choice and male competition in the tettigonia cantans. *Animal Behavior*, **35**:887–900.

Lima, S. L. (1998). Stress and decision making under the risk of predation: Recent developments from behavioral, reproductive, and ecological perspectives. *Advances in the Study of Behavior*, **27**:215–290.

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

Lund, H. H., Webb, B., and Hallam, J. (1997). A robot attracted to the cricket species gryllus bimaculatus. In *Fourth European Conference on Artificial Life*, pages 246–255.

Maes, P. (1990). How to do the right thing. *Connection Science Journal, Special Issue on Hybrid Systems*, **1**:291–323.

McNamara, J. M. and Houston, A. I. (1980). Tha application of statistical decision theory to animal behaviour. *Journal of Theoretical Biology*, **85**:673–690.

McNamara, J. M. and Houston, A. I. (1994). The effect of a change in foraging options on intake rate and predation rate. *American Naturalist*, **144**:978–1000.

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

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 *From Animals to Animats 6: 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*, **56**:1479–1487.

Müller, A. (1925). Über lichtreaktionen von landasseln. *Z. vergl. Physiol.*, **3**:113–144.

Newell, A. and Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, **19**(3):113–126.

Pfeifer, R. and Scheier, C. (1999). *Understanding Intelligence*. MIT Press, Cambridge, MA.

Pirjanian, P. (2000). Multiple objective behavior-based control. *Journal of Robotics and Autonomous Systems*, **31**(1-2):53–60.

Pirjanian, P., Christensen, H., and Fayman, J. (1998). Application of voting to fusion of purposive modules: An experimental investigation. *Journal of Robotics and Autonomous Systems*, **23**(4):253–266.

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.

Russell, S. and Norvig, P. (2003). *Artificial Intelligence. A Modern Approach*. MIT Press, Cambridge, MA, second edition.

Seth, A. K. (2006). The ecology of action selection. *Phil. Trans. Roy. Soc.*

Stephens, D. and Krebs, J. R. (1986). *Foraging Theory*. Princeton University Press, Princeton, NJ.

Thorisson, K. R. (1996). *Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Thorpe, W. (1979). *The Origins and Rise of Ethology*. Preager, New York.

Tinbergen, N. (1950). The hierarchical organisation of nervous mechanisms underlying instinctive behavior. *Sympos. Soc. Exper. Biol.*, **4**:305–312.

Todd, P. M. (1992). Machine intelligence-the animat path to intelligent adaptive behaviour. *IEEE Computer*, **25**(11):78–81.

Tu, X. (1996). *Artificial Animals for Computer Animation: Biomechanics, Locomotion, Perception, and Behavior*. PhD thesis, University of Toronto, Department of Computer Science.

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

von Holst, E. (1935). Über den lichtruchenreflex bei fischen. *Publ. Staz. Zool.*, **15**:143–158.

Werner, G. M. (1994). Using second order neural connection for motivation of behavioral choices. In *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, pages 154–161.