Compromise strategies for action selection

Frederick L. Crabbe*

US Naval Academy, Computer Science Department, 572M Holloway Road, Stop 9F, Annapolis, MD 21402, USA

Among the 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 multiple goals. This paper briefly reviews the history of compromise behaviour and presents 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, and presents some reasons why this may be true and hypothesizes cases where compromise behaviour is truly useful. In particular, it hypothesizes 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’). This paper hypothesizes that compromise behaviour is more beneficial for high- 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 one of those goals, but is good for achieving several of those goals in conjunction. For example, a predator stalking two prey might not move directly towards 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 it 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 the agent’s other 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.

*crabbe@usna.edu

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2007.2053 or via http://www.journals.royalsoc.ac.uk.

One contribution of 15 to a Theme Issue ‘Modelling natural action selection’.

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.

(a) Definitions of compromise

In many approaches to animal behaviour, the full ramifications of an action are weighed in the light of the current situation (see §2d). Since this can be computationally expensive (see §2e), a computational simplification is to divide the action-selection problem into sub-goals, solve those optimally and combine the solutions (see §2b). It is with respect to this latter strategy that compromise behaviour (acting such that no single sub-goal is optimally satisfied) is most often considered (Tyrrell 1993).

Definitions of compromise behaviour can be categorized 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 by 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 not to 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 maximizing the chance of survival or maximizing the chance of reproductive success.

Sections 2b–g 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.

(b) 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 (Timbergen 1950), while the latter considers it synthetically via the construction of agents (Todd 1992; Pfeifer & Scheier 1999).

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; Dewsbury 1978; Lorenz 1981; Brigant 2005), 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 hypothesize 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 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 (e.g. each ear) cause the animal to consider each as a separate goal (Lund et al. 1997). Bimodal input combines signal from two different types of sensors (e.g. eyes and ears) 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 controls 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 towards two light sources (Müller 1925), and in the katydids when performing phonotaxis (Morris et al. 1978; Latimer & Sippel 1987; Bailey et al. 1990). The fish Crenilabrus displays low-level prescriptive, bimodal compromise in its orientation behaviour between its reaction to light and 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 & Cerri 1982), sparrows (Grubb & Greenwald 1982), pike and sticklebacks (Milinski 1986). Indeed, a great many studies suggest that animals balance the risk of predation against foraging or other benefits (Lima 1998; Brown & Kotler 2004).

(c) 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 synthesized (Dewsbury 1992), but early theoretical work had important influences on artificial intelligence (see §2e).

(d) Optimal biological approaches

Modern trends in biology have employed formal models and optimization techniques borrowed from decision theory and operations research (Clemen 1996; Hillier & Lieberman 2002) in order to determine optimal behaviour. Behavioural ecology is the study of interaction between an organism’s environment and its behaviour, as shaped via natural selection (Krebs & Davies 1997). Under the assumption that selection optimizes behaviour to maximize reproductive success, to understand animal behaviour it is important to analyze it with techniques that optimize objective functions which describe reproductive success. For example, in the field of foraging theory (Stephens & Krebs 1986), techniques such as linear programming (Hillier & Lieberman 2002) or dynamic programming (Bertsekas 2005) are used to find optimal foraging behaviour in terms of such features as maximizing energy intake and minimizing exposure to predators (McNamara & Houston 1994; Lima 1998; Brown & Kotler 2004; Houston et al. 2007; Seth 2007). In these studies, optimization is used as a basis of comparison and as an explanation for natural selection; it is not posited as the decision-making process the animal itself
uses. Optimization is computationally expensive; 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 sub-
goals into a single objective function to be optimized.
As such, only the solution to the overall objective
function is considered and not optimal solutions to the
individual sub-goals.

(e) Artificial intelligence and planning

The field of planning within artificial intelligence (AI)
was delayed in development until the advent of robotic
hardware sufficiently sophisticated to exhibit agent-like
behaviour (Fikes & 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 & 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 the
following planning goal:

\[ \text{Have}(\text{robot, medicine003}) \land \text{In}(\text{robot, room342}) \]

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 realised 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
\[ \text{In}(\text{robot, room342}) \]
true while \[ \text{Have}(\text{robot, medicine003}) \]
is false. The robot must then take actions to make
\[ \text{Have}(\text{robot, medicine003}) \]
true which may in turn make \[ \text{In}(\text{robot, room342}) \]
false. This conflict is different from the one between FAPs in that it arises from the order
actions are performed, not from 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 sub-goals 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 sub-goal
in order to improve the quality of others. A survey of
the state of the art in planning can be found in Ghallab
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 &
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.
\[ \neg \text{In}(\text{robot, room342}) \]) are unlike prescriptive goals in
that 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 AI approaches is that attempts
to prove correct actions can be prohibitively expensive
in moderately complex environments. If an agent is
limited to just 10 actions at any time, then each step
into the future increases the number of possible
outcomes to consider by a factor of 10. If the solution
to the current problem is 20 steps long, then the
program must examine in the order of \(10^{20}\) possible
sequences (Russell & Norvig 2003). For comparison,
there have been (an estimated) \(10^4\) seconds since
the beginning of the Universe (Brindle et al. 2003).
This high computation cost prevents the use of these
techniques for planning behaviour with low-level
actions, where solutions to problems might be many
hundreds of steps long.

(f) 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 unforeseen
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 (Brooks 1986,
1997; Arkin 1998). 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.

Since 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 is an important property for
controllers. In 1993, Tyrrell introduced a list of 14
requirements for action-selection mechanisms drawn
from ethology. Of these, requirement 12 was ‘Comprom-
ise 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 characterization is of high-
level compromise. Tyrrell’s list has had significant
impact on the action-selection field (Humphrys 1996;
Decugis & Ferber 1998; Bryson 2000; Girard et al. 2002), and a number of researchers have developed systems to meet the criteria he set out (Maes 1990; Blumberg 1994; Werner 1994; Blumberg et al. 1996; Crabbe & Dyer 1999; Montes-Gonzalez et al. 2000; Avila-Garcia & Canamero 2004).

(g) 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’ (Tyrell 1993), more recent modelling work generated results that seem to contradict the claim (Jones et al. 1999; Bryson 2000; Crabbe 2004); 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- and low-level compromises. 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 who use ethological ideas directly are those who 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 (Thorisson 1996; Tu 1996; de Sevin et al. 2001; Iglesias & Luengo 2005).

Although work mentioned above (Jones et al. 1999; Bryson 2000; Crabbe 2004) implies that compromise behaviour is less useful than originally thought, this work is not conclusive. Section 3 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, i.e. low-level prescriptive and low-level prescriptive. 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; Tyrell 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 prescriptive or prescriptive situations.

(a) 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, 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 maximizes 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 quadrant and the other in the second in the y-range of (0; 100), and in x-ranges of (−100; 0) for one target and (100; 0) for the other. Each pair of targets will be referred to as t1 and t2. The agent can sense the location of each target. Sensor information takes the form of complete knowledge of the locations of the (x, y) coordinates of both targets. Since 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 is in the range (0; 100). The goals will be referred to as G1 and G2. 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 a 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 for the agent to travel that distance. Notationally, d 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 which is not directly towards either target. Any action selected which 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 piecewise linear approximation of a curved path to one of the targets.

(i) 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 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

\[ \text{EU}(A_j|S) = \sum_{S_i \in O} P(S_i|A_j, S) U_h(S), \]  

(3.1)

where \( O \) is the set of possible outcome states; \( P(S_i|A_j, S) \) is the probability of outcome \( S_i \) occurring given that the agent takes action \( A_j \) in state \( S \); and \( U_h(S) \) is the historical utility of outcome \( S \) (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 the least to the greatest, is the same as the order of the corresponding goals. Therefore, EU calculated with utilities is order-isomorphic with EU calculated with goals. For the purposes here, it will be assumed that the goals and utilities are equivalent (\( U(t) = G \)).

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 EU of the actions possible in each state

\[ U_h(S) = U(S) + \max_{A_j \in A} \text{EU}(A_j|S), \]  

(3.2)

where \( A \) is the set of possible actions. The maximum is found because of the assumption that a rational agent will always act to maximize its EU. An agent can calculate EU using multiple actions in the future by recursively applying equations (3.1) and (3.2).

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 (1999). The technique overlays a grid of points on top of the problem space and calculates the maximal EU of each location, given optimal future actions. This is done recursively, starting at the target locations and moving outwards 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 EU of optimal action at an environmental location when the two targets still remain: \( \text{EU}(A_j|t_a, t_b, \lambda) \), where \( \lambda \) is the agent’s location in the environment, \( \theta \) is the angle of the optimal move for the agent, and \( \lambda' \) is one unit away from \( \lambda \) in the direction \( \theta \). By equations (3.1) and (3.2), the EU of being at \( \lambda \) is

\[ \text{EU}(A_j|t_a, t_b, \lambda) = p^2 \text{EU}(A_j|t_a, t_b, \lambda') + p(1-p) \text{EU}(A_j|t_a, \lambda') + p(1-p) \text{EU}(A_j|t_b, \lambda'), \]  

(3.3)

\[ \text{EU}(A_j|t_a, \lambda') = G_a \beta^{l(t_a)}, \]  

(3.4)

and

\[ \text{EU}(A_j|t_b, \lambda') = G_b \beta^{l(t_b)}. \]  

(3.5)

The total EU (equation (3.3)) is the expectation over four possible situations after an action: both targets there; both targets gone; \( t_b \) there but \( t_a \) 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 (3.4) and (3.5). A formal specification of the algorithm is given in the electronic 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 5 and 20 min to converge in these two target scenarios). Instead, many researchers propose easy-to-compute action-selection mechanisms that are intended to approximate the optimal action (Fraenkel & Gunn 1961; Cannings & Orive 1975; McNamara & Houston 1980; Stephens & Krebs 1986; Römer 1993; Hutchinson & Gigerenzer 2005; Houston et al. 2007; Seth 2007). The mechanisms can be divided into two categories: those that select a single target and move directly towards it and those that exhibit some sort of compromise behaviour. In the former category, those considered here are as follows.

— **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 EU 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 towards 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, those examined here are as follows.

— **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(\( \lambda \)) 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 direction of \( V_a \) is AngleTo(\( \lambda, T_a \)) and the magnitude of \( V_a \) is \( G_a/\beta^{l(t_a)} \). The direction the agent moves (\( \theta \)) is

\[ \theta = \text{AngleTo}(\lambda, V_a + V_b). \]

— **Signal gradient (SG).** The agent behaves as if it is following a SG. 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

\[ \theta = \text{AngleTo}(\lambda, \text{arg max}_\lambda \text{EU}(G_a/\beta^{l(t_a)} + G_b/\beta^{l(t_b)})). \]

— **Exponentially weakening forces (EWF).** This strategy is identical to the forces strategy, except the pulling effects of the targets decrease exponentially with distance, rather than quadratically. The magnitudes
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, v. 7.0 and run on a cluster of 25 Sun Blade 1500s, for 347 computer-days. For each scenario, the EU of each of the action-selection mechanisms described in §3a(i) 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 MU strategy). It also reports the percentage of cases where the strategy selected the correct action out of the two possible. MEU is the best of the three as it selects the better target in most cases, and its overall EU is 15% better than MU. MEU selects the worse target only 0.68% of the time. The table also shows that C is a better strategy than MU. This may be so because the EU of a target decreases exponentially with distance, so that closer targets have higher EU 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 over MEU and the percentage improvement over MEU in the single best scenario. There are three important aspects of this table. The 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 by 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.
Compromise strategies for action selection  F. L. Crabbe 1565

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.

(iii) Prescriptive discussion
With respect to animals and natural action selection, the results presented here imply that animals which 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 the two targets. Hutchinson (1999) suggests three possible reasons for what appears to be low-level prescriptive compromise behaviour: (i) a desire not to tip off potential prey that it is being stalked, (ii) it is a part of a strategy to gather more sense data before committing to a target, or (iii) that computational issues yield simple mechanisms that exhibit compromise-style behaviour. Hutchinson’s reasons are particularly interesting in the 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 the value of \( p \). 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 SG or EWF strategies might be the best (although results from foraging theory suggest that animals are able to estimate \( p \) accurately; Stephens & Krebs 1986). Regarding Hutchinson’s third suggestion, Houston et al. (2007) suggest that behavioural characteristics can be ‘side effects’ of rules that evolved in environments which differ from where they are being used, or that the objective function and criteria being maximized 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 hypothesize 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 that natural compromise is occurring. For instance, in the cases of an animal using sub-optimal feeding patches to avoid heightened predator activity, this behaviour could be explained by the animal downgrading the quality of a feeding patch (the \( G_d \)) because of the presence of the predators. The animal then compares the utilities of the two patches directly rather than considering compromise behaviour (Stephens & Krebs 1986). Alternatively, the animal could be abstracting the problem so that it might be solved optimally.

(b) 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 prescriptive 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, p. 170)

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 proscriptive goal to be co-located with the target and a proscriptive goal to avoid being struck by the danger.

(i) Formal model
The model described in §3a 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 timestep, there is a probability \( p_p(\lambda) \) 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 \( \lambda \). The experiments use four different versions of the \( p_p(\lambda) \) function. The agent also has a goal level associated with the target and the danger (\( G_t \) and \( G_d \)), which can vary with the quality of the resource and the damage due to the predator. Other notation remains the same.

The application of equations (3.1) and (3.2) calculates the EU of being at \( \lambda \):

\[
EU(A_\lambda|t, d, \lambda) = p_p(\lambda)p_d(\lambda)EU(A_\lambda|t, d, \lambda') + p_p(\lambda)(1 - p_d(\lambda))EU(A_\lambda|t, \lambda') + p_d(\lambda)(1 - p_p(\lambda))EU(A_\lambda|d, \lambda') + (1 - p_p(\lambda)p_d(\lambda))EU(A_\lambda|d, \lambda'),
\]

(3.6)

and

\[
EU(A_\lambda|t, \lambda) = G_t p_p(\lambda).
\]

(3.7)

The total EU (equation (3.6)) 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

Phl. Trans. R. Soc. B (2007)
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 (3.7). When the target disappears but the danger remains, the agent must flee to a safe distance from the danger, as in equation (3.8). 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 EU strategy, as described in §3a(i). 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 zigzags along the edge of the danger zone as the agent moves towards the target. Active goal is also a non-compromise strategy that acts upon only one goal at a time.

— **Skirt**. This strategy moves directly towards 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 straightaway from the danger.

The EU of each of these non-optimal mechanisms can be calculated for any particular scenario by using equations (3.6–3.8), as in the previous experiments. For these experiments, four \( p_i(\lambda) \) functions were used, all with a danger radius of 20.

— **Linear A**. \( p_a(\lambda) = 0.04 \times (d)^2 + 0.2 \) when \( d \leq 20 \), 1 otherwise.

— **Linear B**. \( p_b(\lambda) = 0.005 \times (d)^2 + 0.9 \) when \( d \leq 20 \), 1 otherwise.

— **Quadratic**. \( p_q(\lambda) = (d)^2/400 \) when \( d \leq 20 \), 1 otherwise.

— **Sigmoid**. \( p_s(\lambda) = 1/(1 + 1.810^{5} - d) \) everywhere.

Linear A was a baseline strategy where the probability of a strike was high near the danger, but low at the low 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.

One thousand scenarios were generated with a target at (50, 90) with \( G_t = 100 \) and a danger at (60, 50) with \( G_d = -100 \). \( p_i \) was varied systematically in the 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 \( p_i \) is too low, the probability that an agent will reach the target quickly approaches zero. Related studies (see Crabbe 2002 and the electronic supplementary material) indicated that compromise behaviour was greater when \( p_i > 0.95 \).) Once the scenario was generated, the EU 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.

(ii) **Proscriptive results**

Figure 2a shows the results of the optimal strategy when \( p_i = 0.995, p_d = 0.99 \) and \( p_o(\lambda) \) 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 maximize 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; Brooks 1986; Arkin 1998). Thus, one would expect it to have no effect on the action selected.

When \( p_d \) is reduced to 0.5, the compromise action in the lower right is less pronounced (figure 2b). The optimal 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 \( p_i \) to 0.95 (with \( p_d = 0.99 \) and \( p_o(\lambda) \) as linear A), the results are qualitatively identical to figure 2a (this and other additional plots can be found in the electronic supplementary material). When \( p_i = 0.95, p_d = 0.5 \) and \( p_o(\lambda) \) is linear A (i.e. low \( p_i \) and low \( p_d \)), predicted compromise behaviour emerges (figure 3a). The combination of 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 nonlinear \( p_o(\lambda) \) functions, compromise action is seen clearly in all cases. Figure 3b shows \( p_o = 0.995, p_d = 0.99 \) and \( p_o(\lambda) \) 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 $p_o(\lambda)$ is quadratic (see the electronic supplementary material).

Comparison between the optimal strategy and the other strategies described above is shown in table 3. The table uses the active goal as a baseline and compares the 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 one-sixth as good inside the danger zone), so was omitted from the table. The percentages are of the average EU for each strategy across all the starting positions and scenarios (200 000 data points). The scenarios are: across all scenarios and starting positions; across just the starting positions that are opposite the target (the lower right region); across the starting positions inside the danger radius; all positions when the $p_o(\lambda)$ is linear A; all positions when the $p_o(\lambda)$ is linear B; all positions when the $p_o(\lambda)$ is quadratic; and all positions when the $p_o(\lambda)$ is sigmoid. Across all samples, the optimal behaviour performs 29.6% better than the active goal, but skirt is nearly as good, performing 29.1% better than the 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 the active goal, but still only slightly so over skirt. This trend continues for locations inside the danger zone and samples from each of the $p_o(\lambda)$ functions.

(iii) Proscriptive discussion

An examination of figures 2 and 3 reveals properties of the optimal strategy that were not initially predicted (see §3b(i)). 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 2b). 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 towards the target (see text).

Figure 2. Optimal behaviour for the agent in proscriptive scenarios. The vectors point in the direction of optimal movement at each given location, and all vectors are of 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 towards the target. (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 $p_o(\lambda)$ is linear A. In most locations, the optimal strategy is to move directly towards 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 without entering the danger zone. (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 towards the target (see text).
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 behaviourally severe, i.e. the $p_n(\lambda)$ was unlikely to generate a strike, so 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 13–18% greater than the active goal strategy that zigzags 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 set-up was based on situations predicted to be amenable towards good compromise actions (Tyrrell 1993), 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 prescriptive cases, it (i) took forms different from what was expected and (ii) 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.

(a) High- versus low-level actions

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

Table 3. Results comparing optimal and skirt strategies to active goal. (The values reflect the average per cent EU improvement of the samples. The rows are subsets of the dataset. The first row is across all the data points (au). The second row is just those points in a region on the side of the danger opposite the target (opp). The third row is those points within the danger zone (DZ). The fourth through seventh rows are for each of the $p_n(\lambda)$ functions (lin A, linear A; lin B, linear B; quad, quadratic; sig, sigmoid), 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 level of 0.999.)

<table>
<thead>
<tr>
<th>scenarios</th>
<th>optimal over active goal (%)</th>
<th>skirt over active goal (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>29.6</td>
<td>29.1</td>
</tr>
<tr>
<td>opp</td>
<td>64.9</td>
<td>63.3</td>
</tr>
<tr>
<td>DZ</td>
<td>26.2</td>
<td>26.1</td>
</tr>
<tr>
<td>lin A</td>
<td>40.9</td>
<td>40.8</td>
</tr>
<tr>
<td>lin B</td>
<td>13.5</td>
<td>13.1</td>
</tr>
<tr>
<td>quad</td>
<td>48.6</td>
<td>48.5</td>
</tr>
<tr>
<td>sig</td>
<td>16.7</td>
<td>15.2</td>
</tr>
</tbody>
</table>
strong. A simple thought experiment 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 \( l_0 \), 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 \( l_0 \) be \( a_\mu \). 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 \( a_3 + b_3 - c(l_3;l_3) > a_2 + b_2 - c(\min(l_0;l_2),l_2,l_2, l_3) \).

Using the council-of-ministers analogy, the \( a \) minister would cast votes for \( l_1 \), but also some for \( l_3 \). Similarly, the \( b \) minister would cast 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 (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 §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 (Gat 1991; Bonasso et al. 1997). 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-based compromise tend to do so only at the behaviour level (Pirjanian et al. 1998; Bryson 2000; Pirjanian 2000).

(b) 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. (i) 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. (ii) In complex environments (e.g. where multiple resources are to be consumed in succession such as the hypothetical scenario depicted in §4a) 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. (iii) In complex environments, optimal or even very good non-optimal low-level actions are prohibitively difficult to calculate, whereas good higher level actions 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 to the idea that while it may exist in nature, low-level compromise behaviour affords little benefit. This paper proposes that compromise is not especially useful at the lower levels, but is useful at higher levels. Future work will revolve around testing, validation or refutation of this Compromise Behaviour Hypothesis.

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, no. N0001404WR20377.

References


Dewsbury, D. A. 1978 What is (was?) the ‘fixed action pattern’? Auk. 95, 310–311. (doi:10.2307/4089344)


