The
University
Of
Sheffield.

Ethological Decision Making with Non-Stationary Inputs Using MSPRT Based Mechanisms

L. F. Nunes

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

2017

Ethological Decision Making with Non-Stationary Inputs Using MSPRT Based Mechanisms

By:

Luana Fernandes Nunes

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Science
Department of Psychology

September 2017

Abstract

One of the most widely implemented models for multi-alternative decision-making is the Multihypothesis Sequential Probability Ratio Test (MSPRT); this model has found application in biological decision-making despite limitations to discrete ('trial-based'), non-time-varying scenarios. Real world situations are continuous and entail stimulus non-stationarity, making them incompatible with the MSPRT; to address this issue, we introduce a new decision mechanism by augmenting the MSPRT with an integration window which allows selection and de-selection of options as their evidence changes dynamically. In this research we explored different scenarios which did not obey the strict derivation of this algorithm, but rather constituted an empirical study of the boundaries of the applications of strict theoretical Sequential Probability Ratio Tests. In order to bridge the abstract laboratory experiments where the inputs tend to be defined as ‘signals’ with the real decisions performed by animals, we took an existing model of shepherd-sheep behaviour and embedded our full MSPRT based model in place of the authors' instant decision mechanism; this allowed the shepherd to perform decisions based on evidence accumulated over time, thus removing the constraint of perfect information being made available to it. The resulting model encompasses the best characteristics of the MSPRT whilst still being representative of ethological decision making algorithms.

Dedication

To my parents, Ana and Fernando, for their love and for being the best parents anyone could wish for, thank you for your tireless support. To my sister Daniela, you're alright!

To Torin, for everything.

To Badger, Cacau, and Caju, and all the friends who kept me sane through the last 4 years.

Acknowledgments

Firstly, I would like to express my gratitude to my supervisor Kevin Gurney who always challenged me to focus my thoughts whilst still seeing the bigger picture. Thank you for always being available to help and for being an inspirational figure; I could not have wished for a better supervisor.

I would also like to thank my fellow PhD students part of the Neuroeconomics research group, as well as James Marshall, Tom Stafford, and Roderich Gross, for the valuable discussions we had over the years which often led me to change my outlook on things.

Finally I would like to extend my gratitude to the University of Sheffield for funding my PhD and providing all the support needed throughout the last four years.

Preface

The work contained in this thesis was conducted by me and is, to the best of my knowledge, original. Part of chapter 2 has been published in: Nunes, L. F. and Gurney, K. (2016). Multi-alternative decision-making with nonstationary inputs. Royal Society Open Science, 3(8):160376.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Decision-Making in autonomous agents

In our daily lives we are continually faced with the problem of ‘deciding what to do next’ based on a plethora of sensory information and our internal cognitive, and homeostatic state. This process of *decision making* continues to attract substantial interest in the computational modelling literature (for reviews see Bogacz et al. (2006) and Usher et al. (2013)). Mechanistically, many of these models assume that noisy sensory streams provide samples of ‘evidence’ for each of several alternatives, and that these evidence samples are accumulated in some way until a reference criterion of distinguishability is reached (defined by one or more thresholds). This is an idea that meshes well with neuroscientific evidence showing ‘evidence accumulation’ in brain areas involved in decision making (Gold and Shadlen, 2007; Heekeren et al., 2008; Churchland and Ditterich, 2012). Such decision making processes are also subject to a speed-accuracy trade-off which accords well with psychological studies of reaction times and error rates (Bogacz et al., 2006). Thus, by manipulating the threshold(s) and integration rates, more or less evidence may be allowed to accumulate before the decision is made. Here, longer accumulation (‘reaction’) times lead to greater accuracy or lower error rates.

1.1.1 Optimality in Decision Making

In the context of decision making, optimality can be defined in a number of ways depending on the task. In contexts where correct decisions are rewarded, the decision threshold usually reflects a compromise between speed and accuracy, with one or the other being favoured in order to maximise the rate of accumulation of reward.

Computational, neural, and numerous other types of models often interpret optimality as statistical optimality, where the aim is to minimise decision time for a fixed error rate (Bogacz, 2007); this interpretation offers a benchmark in which to compare models and is widely accepted as a valid aim for neural models since research suggests that the brain implements statistical tests and therefore has a parallel with statistical decision making models (Bogacz et al., 2006; Bogacz and Gurney, 2007; van Maanen et al., 2012; Lepora and Gurney, 2012; van Ravenzwaaij et al., 2012).

Despite numerous models optimising the trade off between speed and accuracy, in more complex decision contexts where each option is associated with an intrinsic value, the concept of correct and incorrect decision isn't present and therefore models must maximise reward obtained through the value of the options (Kennerley et al., 2006; Pirrone et al., 2014). Pirrone et al. (2014) argue that most naturalistic decisions are value based and that given knowledge about the decision, mechanisms which implement speed-accuracy trade-offs can be parameterised to account for value through the use of value functions.

An example of value-based decisions is nest selection by honeybee swarms; this was described by Reina et al. (2016) as a best-of-N nest-site selection problem, which is equipped with value sensitive parametrisation allowing maximisation of the colony's future fitness.

For the purpose of this thesis, we will consider optimality as statistical optimality where decision time is minimised for a fixed error rate.

1.1.2 Decision-Making between two alternatives

Most previous studies of decision making have focused on the choice between two alternatives (Smith and Vickers, 1988; Ratcliff and Rouder, 1998; Cho et al., 2002; Ratcliff and McKoon, 2008). In order to compare various models and assess how well they can account for empirical data, a simple but widely used method is the two-alternative forced-choice 2AFC task (Feldman and Ballard, 1982). The task consists of the presentation to a subject of spatial or temporal alternatives where a stimulus is shown, the subject then has to choose the location or interval in which the stimulus was presented. This task is extremely useful in testing decision speed and accuracy when deciding between two options with time constraints.

A plethora of models have been developed to account for the data resulting from these tasks, with varying levels of complexity and wide range of parameters, but Cho et al. (2002) found evidence suggesting that simpler mechanisms performed well at fitting the data.

For most decision making scenarios, in order for a correct decision to be made, a decision making mechanism must not evaluate a quantity only at a given point in time but over a period of time. It is then important for the mechanism to be able to store evidence over time; models that use this approach are know as **Sampling Models**.

There are several types of sampling models but for this review we will cover only one particular group, **Accumulator Models** (Vickers, 1970). These models accumulate evidence supporting each Hypothesis, and compute it in a principled way, stopping when a given threshold is reached.

The simplest type of accumulator models are the Race models; these have separate accumulators for each hypothesis and accumulate evidence supporting each of them by summing the inputs to the model. The integration stops as soon as one of the accumulators reaches a threshold (Smith and Vickers, 1988), and the hypothesis corresponding to this accumulator is considered correct.

While this type of model might result in quicker decisions than other accumulator models, it is inefficient because it does not take into consideration the difference between the alternatives. In order to overcome this shortcoming, many models with increased complexity have been developed over the years.

One widely used model which is arguably more efficient than the Race Models is the Drift Diffusion Model (DDM) (Ratcliff et al., 1978; Busemeyer and Townsend, 1993; Ratcliff and Rouder, 1998; Gold and Shadlen, 2001; Roe et al., 2001; Ratcliff et al., 2003). Like the race model, the DDM accumulates evidence over time for each of two competing options but in this case, using only one accumulator.

The reason for the use of a single accumulator is that, effectively, evidence supporting one option is in turn evidence against another, and so the model employs not one but two decision thresholds - or boundaries. In the DDM model, the boundaries are located a certain distance above and below the starting point of evidence accumulation and whichever option is stronger will bias the evidence in the direction of one of the two boundaries; since the model has two symmetric bounds (a positive and a negative mirrored bound), defining μ as the mean evidence value, if $\mu > 0$ evidence supports hypothesis 1 and if $\mu < 0$ evidence supports hypothesis 2. The diffusion term can be interpreted as the noise term and is obtained from a Gaussian distribution of mean 0.

The DDM is significantly different from more basic accumulator models due to an added parameter, the drift rate (Ratcliff and McKoon, 2008). This parameter determines the rate at which evidence is accumulated and by doing so limits the speed at which any given option can reach the boundary and get selected. Because it is directly connected with the decision time, the drift rate will vary according to the difficulty of decision at hand - two closely matched options should result in a more difficult decision and therefore need more samples of evidence. This parameter reflects the balance between decision time and desired accuracy (Ratcliff and McKoon, 2008; Ratcliff and Rouder, 2000).

In contexts where the stimulus statistics can vary, the practical applications of the DDM are limited, since its drift rate is a ‘stationary’ parameter which depends on these statistics and therefore requires external input. Another shortcoming of this model is that even though there is a single accumulator for the two competing options, there is no inhibition between them, so that an increase in accumulated evidence can not speed up the decision process and therefore will result in a sub-optimal decision time (not the minimum decision time for a given error rate). Recently, Bitzer et al. (2014) showed the equivalence between the DDM and Bayesian models of decision making under certain conditions and how it might be implemented in the brain.

In 2001, Usher and McClelland proposed a model with added complexity and which addressed some of the shortcomings of the DDM, the Leaky Competing Accumulator model (Usher and McClelland, 2001; Ratcliff and Smith, Philip, 2004; Smith and Ratcliff, 2004). The Leaky Competing Accumulator model (LCA) as described by Usher and McClelland (2001) is a decision making model that, as the name suggests, uses two competing accumulators that inhibit each other while at the same time having intrinsic leakiness. The authors proposed that the combination of leak and activation made the model more biologically plausible, distinguishing it from others such as the DDM.

The DDM is equipped with two free parameters which change the activation of the accumulators as time progresses. The first of the two parameter is leak, and it represents the proportion of accumulated information that is lost over time. This is constant across accumulators and means that even thought the proportion of loss will be the same, the actual loss over time will depend on the overall amount of accumulated evidence.

The second free parameter is inhibition, and in this particular case, lateral inhibition. As the amount of evidence grows in one accumulator, the inhibition from that to the other accumulator also grows, and so the more evidence is accrued by a particular accumulator, the more strongly it will inhibit the other, causing it to decrease its overall accumulated evidence. A situation where one option is clearly superior can be easily identified from the evidence accumulated for the other since it leads to a more pronounced decay of evidence over time. The use of inhibition in the LCA allows it to operate with only one fixed decision boundary since the overall amount of accumulated evidence already reflects the competition between options. Since the combination of the two parameters dictates the rate of accumulation of evidence and consequently the decision time, this can be thought of as being equivalent to the drift parameter in the DDM, although the drift in the LCA, as previously noted, incorporates an extra component of inhibition.

There is a large division in the literature about which mechanism is considered optimal and under which circumstances. Bogacz et al. (2006) state that models like the LCA can become optimal by approximating the DDM, while van Ravenzwaaij et al. (2012) argue that this approximation is lost under certain circumstances and so is the optimality. Overall, there is no one optimal decision making model over all types of decision, since there is a large variability in conditions, which make certain models optimal under certain circumstances but not others.

Recent work by van Ravenzwaaij et al. (2012) scrutinising the model equations proposed by Bogacz et al. (2007) for the LCA model showed some discrepancies in its formulation. One of them is that the equations required negative activation, which in turn requires negative firing rates. If these rates were truncated to 0 to avoid this, the equations ceased being valid. The second discrepancy was the trial-by-trial variation of the boundary separation parameter which breaks the equality between LCA and DDM. However, by assuming the existence of baseline activity for each accumulator, the authors were able to approximate the DDM results with the LCA model. Despite that, under the evidence presented before, both models cannot be considered identical. In this case, the question still holds on whether biologically plausible decision making models can, at the same time, be optimal.

For the case of the 2-AFC, a widely used task to test decision models, the DDM is considered to be optimal (Laming, 1968; Bogacz et al., 2006) for yielding the fastest possible decision time for a fixed accuracy. This model works in continuous time and so is significantly different from discrete sampling models. When the frequency of samples is reduced, these can no longer approach continuous sampling and the optimality of the DDM is no longer maintained. In this case, an alternative algorithm is available, which is the discrete counterpart of the DDM, the Sequential Probability Ratio Test (SPRT) (Smith and Vickers, 1988). In this regard, the SPRT has been shown to be optimal in that it takes the least number of evidence samples to reach a decision for a given error rate (Wald and Wolfowitz, 1948).

The SPRT algorithm works by establishing two hypotheses, one supporting each alternative, which are associated with probability distributions. The sampling of those distributions will lead to the calculation of log-likelihood ratios for each hypothesis. These ratios will be accumulated until one of them reaches the threshold, dictating the selection of the hypothesis associated with it as true and terminating the decision process.

This algorithm has also been successful at accounting for behavioural data (Ratcliff et al., 1978; Schall, 2001; Ratcliff et al., 1999), with neuronal recordings presenting a similar working structure to it. These recordings show that groups of neurons linked with a response increase their firing rate to a threshold in order to trigger that response, in the same way that the SPRT does.

In the LCA, when the inhibition and leak terms are equal, Bogacz et al. (2007) showed that the model is equivalent to a DDM which in turn means it can implement the SPRT. This characteristic, like in the previous models, makes it possible for the LCA to achieve both optimality, as described, and biological plausibility.

1.1.3 Decision-Making between multiple alternatives

Despite the advantages associated with the use of each the previously mentioned algorithms, in their current form, these can only be utilised in a two alternative decision making scenario. Recently, however, there has been a growing interest in understanding the processes involved in multi-alternative decision making (more than two options) (Churchland et al., 2008; Churchland and Ditterich, 2012).

In 2008, Churchland et al. (2008) directly compared the responses of trained monkeys in a four-choice against a two-choice discrimination task and found that in the multi-alternative task, a higher decision threshold was used, as indicated by the lower firing rates at the beginning of the trial; according to the authors, when the number of alternatives is increased, uncertainty is also increased. These results suggest the need for more evidence to be accumulated in tasks with increased number of alternatives in order for the threshold to be reached, which is indicative of an increased difficulty associated with said tasks.

Another interesting effect of increasing the number of alternatives is a phenomenon known as preference reversal (Roe et al., 2001). A typical example is when deciding between alternatives 1 and 2, if the subject has chosen 1, the addition of a third lower value alternative can shift the decision to 2. This is because in multi-alternative decisions the context also affects the choice (Churchland and Ditterich, 2012).

This phenomenon was analysed and simulated in greater depth by Tsetsos et al. (2010), who looked at the three types of reversal effects: attraction, compromise and similarity. Simulations of Decision Field models and LCA indicated that the second was the best at predicting these effects, perhaps due to its development aimed at fitting perceptual choice tasks. As such, models of decision making that accommodate multi-alternatives have to explain the previous effects in order to be deemed biologically plausible.

In recent years, research has also focused on extending models originally developed for binary decision to multi-alternative computation (McMillen and Holmes, 2005; Ditterich, 2010; Krajbich and Rangel, 2011).

One of such extensions was proposed by Krajbich and Rangel (2011); in their work, the authors took known findings about binary-choice and generalised them to ternary choice models. This extension used the parameters estimated for binary choice and proved accurate at describing the relationship between choice and reaction time data for three alternatives. The authors argued that the computational processes involved in binary choice were similar to those involved in trinary choice.

Despite the clear need for multi-alternative decision models stemming from the search for ethological plausibility, the number of models which have currently been extended to perform multi-alternative computations, as well as models developed exclusively for that purpose is limited.

Perhaps one of the most widely implemented models for multi-alternative decision making is the *Multihypothesis Sequential Probability Ratio Test* (MSPRT) (Baum and Veeravalli, 1994; Dragalin et al., 1999, 2000; McMillen and Holmes, 2005). While this reduces to SPRT for two alternatives, it fails, in general, to reach similar optimality criteria. Thus, the MSPRT can be shown only to achieve what is known as 'asymptotic optimality' in that it minimises mean response time as the error rate tends to zero. Even though optimality is not achieved if error rates are not negligible, the performance of MSPRT in this case is still better than some of its competitor mechanisms. Furthermore, there is no known mechanism which is optimal for all levels of desired accuracy (Ditterich, 2010).

The DDM, like the SPRT, was developed originally as two alternative decision model with the computations necessary for its extension to multi-alternative unknown until recently. In 2011, Krajbich and Rangel (2011) proposed a model capable of performing ternary choices. This implementation was very closely linked to the two alternative one with the parameters fitted to the binary test still fitting the results of this extended version. One of the most important findings by the authors was that if it is plausible for the implementation of this algorithm to occur in the brain, then the processes used by it in the two types of choices are similar.

Authors chose the MSPRT framework in this extended model since it could describe the experimental data extremely accurately. This also allowed the model to produce asymptotically optimal computations in the absence of attention biases. Despite that, the model is more of an MSPRT-like mechanism that an exact implementation of it.

Interesting experimental results, accurately described by the model, showed that when making a decision, the subjects would generally take all options into consideration, despite fixations toward the end of the trial focusing mostly on the best option. An interesting point made by the authors is that their implementation of a multi-alternative DDM is not necessarily the best, this is because there are multiple implementations of the DDM in the literature, meaning also that model comparison is more challenging. Another good point made by McMillen and Holmes (2005) is whether or not optimality is desirable in every scenario.

Like the multi-alternative implementation of the DDM, the one for the LCA model is also easily generalised from its two choice counterpart. There are however, many ways in which this could be implemented depending on the assumptions made about the leak value, bounds and the actual implementation of the integrator(Churchland and Ditterich, 2012).

As argued previously, the LCA model can not be considered identical to the DDM model, which in turn results in it not being able to strictly implement the MSPRT. When applied to multiple alternatives, the LCA model approximates a statistical test different from the MSPRT, a suboptimal statistical test known as the max-vs-average test (McMillen and Holmes, 2005); despite that, Ditterich (2010) showed that the optimality advantages of the MSPRT were greatly reduced for error rates of 20%; in such cases, the use of a max-vs-average test which is much less computationally expensive to implement might be sufficient.

1.1.4 Basal Ganglia and the MSPRT

Given the SPRTs links with behavioural data (and it's extrapolation into mutiple competing channels through the MSPRT) and building from the premise that the basal ganglia acts as a central switch for action selection (Redgrave et al., 1999), and that the computations are statistically optimal; Bogacz and Gurney (2007) have proposed that this structure has the required anatomy and physiology to implement the MSPRT.

The basal ganglia are a set of evolutionary old, subcortical nuclei which are believed to play a critical role in action selection and decision making (Redgrave et al., 1999; Ding and Gold, 2013). In recent decades, a number of studies have shown the involvement of the basal ganglia and certain cortical regions in action selection (Chevalier et al., 1985; Chevalier and Deniau, 1990; Shadlen et al., 2001).

Due to the large number of connections between the basal ganglia and other structures in the body and the fact that it can process large amounts of information, Redgrave et al. (1999) proposed the idea of this brain region acting as a central selection device. This idea was further developed in 2001 with (Gurney et al., 2001b,a), where the authors presented a computational model that encoded evidence supporting each possible action as a measure of salience (scalar), representing the overall activity level of the neural representation of that action. It was also proposed that the medium spiny neurons associated with the action were the structures responsible for the extraction of this salience.

The mediation of actions was accomplished through the release of inhibition with two possible mechanisms: *hard switching* - only one input can be selected; *soft switching* - more than one input can be selected.

Due to the noisy nature of stimulus input, the evidence supporting each channel can be thought of as having statistical nature (Wiesenfeld and Moss, 1995; Gold and Shadlen, 2001). For that reason, and assuming the brain has evolved in a way that will optimise the decision process, it is likely that it does implement statistical tests (Bogacz, 2007).

In 1948, Wald and Wolfowitz (1948) showed that the Sequential Probability Ratio Test $SPRT$ was the statistically optimal decision making algorithm by proving that it required fewest observation to reach a decision (Wald and Wolfowitz, 1948). The mapping of MSPRT to basal ganglia anatomy requires the existence of precise anatomical and physiological aspects present in the the basal ganglia (Bogacz and Gurney, 2007), supporting the view that these nuclei may be at least approximating the decision algorithm under certain circumstances. While we are not interested in studying the basal ganglia's role in action selection as such, we are interested in maintaining the link between this structure and the MSPRT.

1.1.5 Non-Stationary Decision-Making

Irrespective of mechanism studied, or the number of choices involved, most work on decision making thus far has assumed a discrete *trial based* paradigm in which evidence is accumulated from streams with stationary statistics, a decision reached, and all state variables are reset. Usually, in trial-based schemes, the decision is made when the threshold criterion is reached but, in the *interrogation paradigm* (Bogacz et al., 2006), a decision is forced after a predetermined time period. In either case, statistical stationarity and mechanism reset is assumed before the following trial period.

Recently, the ethological validity of these assumption has been questioned by considering non-stationary stimuli (Ossmy et al., 2013; Tsetsos et al., 2011); however, these studies still assume trial-based schemes, and also use the leaky competing accumulator decision mechanism (rather than the MSPRT). In a fully ethological situation, we expect that, as well as stimulus non-stationarity, there will be no clearly delineated 'trials', implying no obvious way of resetting the decision mechanism, and thereby *de-selecting* the current choice.

Certain neural decision making algorithms based on probabilistic algorithms have a built in way of dealing with evidence weighing and discarding. Such mechanisms, like the LCA, are equipped with parameters like leak, or even weighing parameters, which allow for older, and arguably less relevant evidence to have its weight in current decisions lowered or for it to be discarded altogether.

In Tsetsos et al. (2012), the authors showed that the LCA was capable of accounting for different patterns of behaviour from human data, due to the fact that it did not assume a temporal bound for decision; in this paper, the authors explored the use of non-stationary evidence in testing decision models; for this, variations of the race and diffusion models and the LCA model were employed, which were capable of performing multi-alternative choice. The pattern of the inputs used was relevant to our work since it did not follow a stationary pattern of inputs, therefore presenting the added variability we were interested in exploring.

The MSPRT model was tested with these inputs and results showed effects which were consistent with the basal ganglia model, in which a 'selection limiting' effect was observed. This selection limiting effect relates to the delay in selection that may incur from having two closely matched options instead of two significantly different ones.

The question we sought to address was, therefore: how can a decision mechanism continually select *and* de-select actions or choices governed by multiple streams of noisy sensory information? In a trial-based, reset-endowed scheme, 'old' evidence from the previous trial is explicitly discarded (at reset) before a new decision is made. In the proposed paradigm, we have to implement an automatic, dynamic mechanism for 'forgetting' old evidence.

1.2 Real world applications

After establishing the robustness of MSPRT-like mechanisms to variations in input dynamics and showing that the theoretical constraints can be broken while retaining some of the qualities which make the Sequential Probability Ratio Test an extremely popular model of decision making (optimising the balance between accuracy and decision time); we were interested in exploring real world applications of our algorithm, where the inputs were no longer viewed as Gaussian signals but rather a result of the behaviour of external agents. These agents could be anything from people to sheep and would be equipped with a set of behaviours particular to the group and potentially the individual.

By embedding a computational decision model into a shepherd agent, an ideal testing scenario can be achieved since the agents to be herded can be used to model inputs which cannot be predicted by the shepherd or engineered to match the constrained laboratory scenarios; this unpredictability stems from the fact that the shepherding agent can only, at best, estimate the positions, speeds, and directions of movements of the agents, and therefore the inputs to the model will always include sensory noise.

In the last decade, a lot of interest has arisen in understanding the behaviour of collective agents, resulting in an extensive body of literature looking at the various aspects of group behaviour, from self-organisation (Helbing et al., 2001, 2005; Cho et al., 2007; Moussaïd et al., 2009) to crowd dynamics (Langston et al., 2006; Singh et al., 2009; Cristiani et al., 2011; Mehner et al., 2015; Bouchard et al., 2015; Aggarwal and Goatin, 2016).

One area which has attracted a lot of interest in recent years is that of flocking and herding of groups of agents (Miki and Nakamura, 2006, 2007; Bennett and Trafankowski, 2012; Lakshika et al., 2015; Brulé et al., 2016; Gade et al., 2016).

A lot of different algorithms have been proposed to model the behavioural patterns of shepherding as well as flocking agents, but despite the wide range of implementations of shepherding dynamics, the goal of all models is to move the flock of agents to a target in a limited amount of time.

1.2.1 The shepherding problem

Over the years, many algorithms modelling shepherd behaviour have been proposed, many of which are fundamentally different in operation. The simplest of those tend to model only the sheep-sheepdog interactions, without any outside agents controlling the sheepdog's behaviour.

One of such models was described in Miki and Nakamura (2007), in which one or more shepherds were used to control a flock of agents. The shepherd's behaviour followed four simple rules: guidance, flock making, keeping, and cooperation, which were intended to move the flock in a particular direction, collect wandering individuals, avoid breaking flock order by keeping a minimum distance from agent, and avoid collision with other shepherds, respectively.

The shepherd or shepherds made use of several types of sensors in order to obtain information and estimate the position of each individual agent. Authors found the proposed model to be effective when herding up to 25 sheep agents with the use of either a single shepherd or two shepherds working in cooperation.

One model which requires more than one shepherd was proposed by Jyh-Ming Lien et al. (2005); in their work, the authors focused on optimising herding strategies for multiple shepherds working in cooperation. Despite the shepherds having to work cooperatively, the authors focused on strategies which allowed this cooperation to take place without communication between shepherds.

One of the reasons for choosing multiple shepherds over a single one was that their initial experiments indicated that as the flock size is increased significantly, a single agent is no longer able to adequately control the flocking agents.

The authors applied their algorithm to a range of flock dynamics with different levels of separation. This separation reflected each flock's tendency to group or scatter with the approach of the shepherd. Different groups of animals have different separation tendencies with flocks of sheep tending to present lower separation than flocks of cattle or ducks. The single shepherd also failed when trying to control flocks with higher separation tendencies, with multiple shepherds performing significantly better.

Alternative models to those described previously follow a shepherd-sheepdog pattern, in which, effectively, two different agents are needed to herd the flock to the target: a shepherd, and a sheepdog. In this case, it is the shepherd's task to communicate a set of commands to the sheepdog in order to get it to herd the sheep agents to the target.

Bennett and Trafankowski (2012) proposed a shepherding mechanism with this pattern, which incorporated commands inspired by those used by real shepherds when communicating with sheepdogs. The authors modelled what is known as circling behaviour, in which the dog switches between clockwise and anticlockwise motion around the flock, thereby forcing the sheep to stay grouped while still moving in the desired direction, away from the shepherd.

Due to the behavioural patterns exhibited by the sheep, the sheepdog must maintain this circular motion while at the same time moving towards the desired target. It is the balance between these two different motions which dictates the success of the shepherding event. This algorithm was tested across a wide range of flocking characteristics and found effective, although it proved very sensitive to parameters determining changes in direction of the sheepdog movement.

With the quick development, in recent years, of the game industry and the use of artificial intelligence to model characters and populations, shepherding algorithms have also been developed with such applications in mind. Cowling and Gmeinwieser (2010) modelled a shepherd agent using common artificial intelligence techniques in a two dimensional game world. Their implementation was comprised of two separate shepherd entities, one who issues commands, which in a gaming scenario would be replaced by the human player - the shepherd - and one who obeys the commands - the sheep-dog.

The authors built on the work of Jyh-Ming Lien et al. (2005), but focused on how these behavioural patterns could be implemented in a game environment. Their implementation proved a challenge to human users, making the game engaging and interesting.

Recent work by Brulé et al. (2016) differed from previous research in that it applied Genetic programming algorithms to implement the sheep-dog agent. Their algorithm worked with either a single dog or multiple dogs, with cooperative herding. The main difference between this implementations and previous work is that the sheep-dog agent was implemented as a memoryless function acting on a set of parameters passed by the simulation. The genetic program would then evolve to produce behaviour which resulted in improved fitness. Fitness was defined by the authors as the percentage of sheep that were successfully shepherded into the target area.

The increase in number of generations would result, subject to trial-by-trial variations, in growth in fitness, converging to close to perfect shepherding after a number of generations. The resulting algorithm would evolve to create shepherding strategies which were adapted to the herd.

All of the research described before has focused on modelling the shepherding agent or agents, and in using different methods to improve the number of successful shepherding events. However, something just as important as the shepherd and that ultimately influences the success of failure of an event are the heard dynamics. From sheep to cattle, different groups of animals or robotic agents will exhibit different group behaviours.

Certain types of flocks are more prone to clustering together when faced with perceived danger, whereas other will tend to scatter more. If we think of a group of people as our flock and we confront them with a situations which is perceived as dangerous, we know that due to our complex patterns of behaviour and the fact that we are rational beings, we are more likely to stay in the proximity of those who we know better, be it a friend, colleague or family member. In this case, in a situation of fire, for example, we know that most people would attempt, if possible, to find those people and not just cluster together with those in their vicinity .

Due to these different group dynamics, much like with any other type of decision making, it is difficult to find a shepherding model which is optimal when applied to all possible flocks.

Generally, algorithms capable of adapting or evolving according to environmental conditions (Brulé et al., 2016) tend to perform better in a wider range of scenarios.

1.2.2 Flock dynamics

One parameter common to most flock dynamics is 'flight distance' (Evered et al., 2014), this is used to describe the maximum distance at which an element of the flock will perceive the shepherd as a threat and engage in a set of self-preservation behaviours like moving away from the shepherd, and/or closer to other elements of the flock. This distance will vary for different group dynamics and situations.

From research conducted into the subject, Evered et al. (2014) found that, unlike what is assumed by most models of group behaviour, this 'flight distance' parameter is not constant, and was actually found to decrease as the trial progressed, as well as from trial to trial, as the flock became more accustomed to the robotic agent. Contradicting results were presented by King et al. (2012), who found that flight distance actually increased from the first trial. This was attributed to the fact that unlike Evered et al. (2014), who used a robotic agent, King et al. (2012) used a real sheepdog. These results show that extrapolating results obtained from simulations with robotic agents to real agents must be done with caution since the two are perceived different by the flock.

Despite flock-by-flock variations in behaviour, there are usually three main behaviours which all flocks tend to exhibit : clustering, avoidance of shepherd, and avoidance of barriers (Cowling and Gmeinwieser, 2010; Bennett and Trafankowski, 2012; King et al., 2012; Evered et al., 2014; Brulé et al., 2016).

The first behaviour, clustering, is built from the balance of attraction and repulsion from other agents. In this case, depending on the distance between two agents, they will either exhibit higher attraction or repulsion. Generally speaking, the closer the agents are, the more strongly they will repel each other, and the further away, the more strongly they will attract. This makes sense from a self-preservation perspective since in a situation of danger, it is more advantageous for the agents to stick together. This balance of attraction-repulsion will depend on the particular group and will determine the herd's tendency for separation.

The second behaviour exhibited by the herd is avoidance of shepherding agent. Since one of the components of shepherding is the need for the herd to perceive the shepherd as posing some sort of danger, the avoidance behaviour, like the clustering, stems from a self-preservation attitude in which the agents perceive the shepherd as dangerous and as such will attempt to keep a safe distance whenever possible.

The third behaviour, barrier avoidance, is self-explanatory and is just used to guarantee that in a simulated environment the sheep stay within the limits of the simulation and also that they will try to avoid being stuck to the sidelines. This is important when the shepherd's position pushes the sheep toward the sidelines, so that they don't all get stuck in a deadlock situation.

Despite the large number of available models for shepherding, the majority of those are based on general assumptions of herd behaviour, like the three components presented previously. More realistic models, which are mode likely to translate back into real world applications, should model herd dynamics in a more detailed way, and take into consideration the wider range of individual and group behaviours exhibited by the agents.

Strömbom et al. (2014) presented one of the more realistic models available, by modelling sheep and sheepdog behaviour according to a self-propelled particle model of local attraction-repulsion. This implementation reproduced key features of empirical data collected by the authors from real life sheep-sheepdog interactions. In it, the shepherd is faced with N sheep in random positions, which it has to herd to a target. At each time step the shepherd is faced with the decision between herding all sheep to the target or collecting a particular outlier, based on the sheep's positions.

Due to their arguably more realistic implementation of the herd, we believe that this algorithm can generate realistic input saliences and therefore constitutes an ideal scenario to test the extended MSPRT's decision making capabilities.

Chapter 2

MSPRT algorithm with non-stationary evidence

2.1 Extending the MSPRT to accommodate non-stationary evidence

In the MSPRT algorithm as described by Bogacz and Gurney (2007), the means are hypothesised to follow a pattern where N channels are competing, with N-1 losing and 1 winning channel. The winning channel is defined as that whose mean is μ_1, while all losing channels will have a mean of μ_2; with $\mu_1 > \mu_2$

In this scenario, all losing channels have the same input mean resulting in a constant difference between the means of the winning channel and any of the losing ones. The signal pattern of the competing channels is that in fig. 2.1, where the winning channel is shown in red, and the losing channels are shown in green and blue.

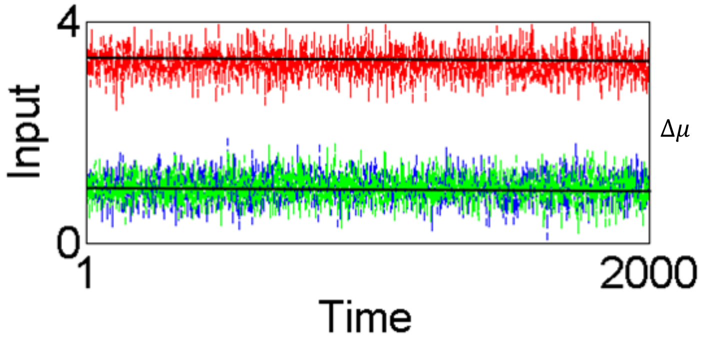


Figure 2.1: Constant input pattern with two phases; each phase corresponds to a fixed mean. Blue and green signals have mean μ_2 and red signal has mean μ_1; $\mu_1 > \mu_2$.

In a more ethologically plausible scenario, one would expect to see more variability in the inputs, where some or all input channels could have different means that would change over time – time-varying.

Tsetsos et al. (2011) shed some light on the effect of time-varying inputs on selection by using a phase transition protocol, for this, the authors tested several known decision making models, and examined how well these could account for the preference reversal effect observed in human subjects. The goal of the experiment was to verify whether or not the closely matched competing means contribute positively or negatively to the selection, for a range of decision thresholds.

We tested the MSPRT in the same conditions and found that the correlation of options can contribute positively or negatively to the selection.

Since we were interested in time-varying inputs and running the algorithm continuously, we sought out to implement an automatic mechanism capable of 'forgetting' outdated evidence, taking as our starting point the MSPRT algorithm. This was done in order to allow subsequent contact with possible basal ganglia implementation, and benefit from the potential for approximate optimality offered by this mechanism.

In our extension to the MSPRT algorithm, we started with a conceptually easy mechanism of temporal sampling, a rectangular sampling window. This window had a fixed duration and worked by performing a convolution of the evidence samples with a rectangular function $w(t)$ of duration t_w.

In order to equip the algorithm with another level of abstraction from human input we devised a mechanism which, based on the difficulty of the decision at hand (difference between means of competing signals), dynamically selected a window size so that more difficult problems had larger windows, allowing them more time to accumulate evidence; conversely easier problems had shorter windows since less evidence was needed to reach the threshold.

The main shortcoming of this implementation was the fact that due to the varying nature of the window size, each evidence sample had to be stored in order to allow for the use of more or less evidence as required, which carried higher costs in terms of memory requirements than algorithms such as the Leaky Competing Accumulator (LCA), which accumulates evidence supporting each alternative using separate accumulators (Usher and McClelland, 2001) and requires only that the sum of inputs be stored. The LCA model is equipped with two terms which result in the decrease of accumulated evidence, a leak term and and an inhibition term; the inhibition term is a result of the competition between options and the leak denotes the rate of evidence decay; this reduces to a simple race model when both terms are 0.

To address the high memory requirements and in line with other accumulator models, we devised a second type of window that no longer had rectangular geometry, but rather exponentially decayed the weight of the evidence. This mechanism worked by continuously summing input evidence and storing only the accumulated sum. The accumulated sum was then multiplied by a constant factor at each time step, which resulted in an exponential decay of evidence samples over time.

The λ parameter was generated from the same function which outputted the window size for the rectangular window, so that $\lambda = 1/DT(\Delta\mu)$, with $DT(\Delta\mu)$ corresponding to the rectangular window size.

The constant factor was derived in a similar matter to our rectangular window size so as to allow consistency in time scales across mechanisms.

This mechanism resulted in lower levels of reward than its rectangular counterpart, but presents a suitable alternative for implementations with more memory restrictions.

2.1.1 The Multihypothesis Sequential Probability Ratio Test

We describe here our starting point – the MSPRT algorithm, conceived as a trial-based process, and interpreted in a form suitable for mapping to basal ganglia operation.

Consider N streams or *channels* of sensory information $x_i(t)$, which are samples at discrete times t, from stationary random variables (fixed mean and standard deviation).

Forming a hypothesis H_i, for each channel i, which is that $x_i(t)$ is sampled from a distribution f^+ with mean and standard deviation μ^+, σ, while all the others are sampled from counterparts f^- with parameters μ^-, σ, where $\mu^+ > \mu^-$. If $X(T)$ is the entirety of sensory information up to some time T, we wish to find posterior probability $P(H_i|X(T))$. It can be shown (Bogacz and Gurney, 2007; Lepora et al., 2012) that the required input to a Bayesian decision mechanism for determining this are the 'evidence' samples formed from the per-sample log-likelihood ratios

$$LR_i(t) = \ln \frac{f^+(x_i(t))}{f^-(x_i(t))} \tag{2.1}$$

Since our research is focused on the use of Gaussian random variables, we now specialise the algorithm to these variables and assume their use from this point forward. In this instance, (2.1) becomes

$$LR_i(t) = gx_i(t) - g\frac{\mu^+ + \mu^-}{2} \tag{2.2}$$

where

$$g = \frac{\mu^+ - \mu^-}{\sigma^2} \tag{2.3}$$

It transpires that channel independent terms do not affect the posterior, and so we may ignore the second term in (2.2) and use $LR_i(t) = gx_i(t)$. That is, we can take as our evidence samples the original sensory samples themselves, $x_i(t)$, interpreting g as an overall 'gain' factor. This reduction is a special case applicable only to Gaussian random variables; other distributions yield more complex expression (Caballero et al., 2015). However, given Gaussian noise, the MSPRT has the interpretation of effectively finding the input channel with the largest mean of its stimulus samples $x_i(t)$; an interpretation we will use henceforth. The accumulated evidence, $y_i(T)$, up to time T, is then just

$$y_i(T) = g \sum_{t=1}^{T} x_i(t) \tag{2.4}$$

Using Bayes theorem, it may now be shown (Bogacz and Gurney, 2007; Lepora et al., 2012) that the log of the posterior $L_i(T) \equiv \ln P(H_i|X(T))$ is given by

$$\begin{aligned} & L_i(T) = y_i(T) - Z_i(T) \qquad (2.5) \\ \text{with} \qquad & Z_i(T) = \ln \left(\sum_{k=1}^{N} \exp(y_k(T)) \right) \end{aligned}$$

A decision is reached when $L_i(T)$ crosses a threshold θ from below. However, the basal ganglia work by release of inhibition from their target structures in the brain (Chevalier and Deniau, 1990); that is 'winning' channels are associated with a decrease in signal values. Thus, if we wish to interpret the output of the MSPRT as the output of the basal ganglia, we need to consider not $L_i(T)$ but $\bar{L}_i(T) \equiv -L_i(T)$, so that

$$\bar{L}_i(T) = -y_i(T) + Z(T) \tag{2.6}$$

Now, a decision is reached when $\bar{L}_i(T)$ crosses the threshold from above (Bogacz and Gurney, 2007), which is the scheme we adopt here.

The term $Z_i(T)$ captures the interaction between channels; for a given amount of evidence $y_i(T)$ for channel i, increasing evidence for other channels increases $Z_i(T)$, thereby tending to prevent a decision. We therefore refer to $Z_i(T)$as the *conflict term*.

2.1.2 Testing the algorithm with phase-shifting inputs

In order to test whether having closely matched competing means increases the difficulty of selection of an option, we implemented the protocol described by Tsetsos et al. (2011), using the MSPRT algorithm.

The input protocol is shown in fig. 2.2, where the competing channels represented by the colours blue, green and red are A,B, and C, respectively. There are two possible phases, with phase one following the pattern: $\mu_A = \mu_B = 0.8, \mu_C = 0.4$, and phase two following the pattern: $\mu_A = \mu_B = 0.4, \mu_C = 0.8$. The transition between the two phases within a trial is determined by a Markov process, with each trial starting at a random phase; the probability of switching phases increases with the number of time steps since the last switch (n), with $p(n) = 5 \times 10^5 \times n$; the length of each trial is uniformly chosen from 5-10s. An example of a task with the protocol described would be one where the participant had to identify which of a set of three lights of different colours is the brightest, with this brightness changing at intervals dictated by the Markov processed described previously; further, at any point in time, the blue and green lights have the same brightness, either higher or lower than that of the red light.

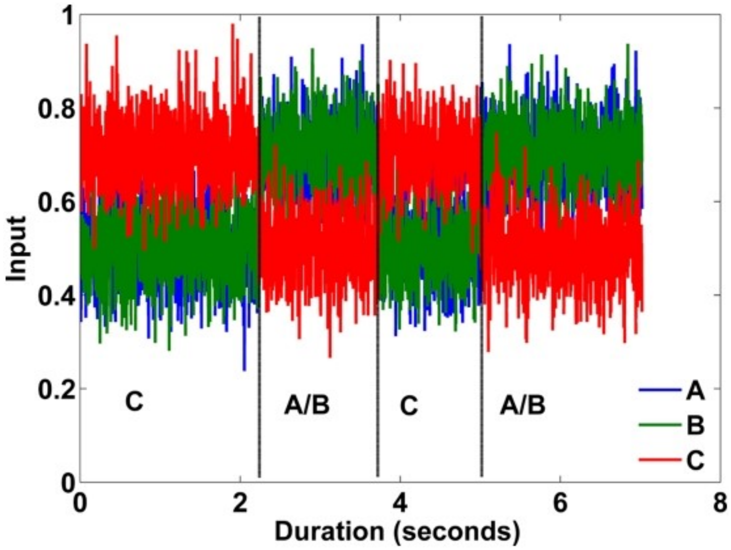


Figure 2.2: Evidence for each competing channel (Tsetsos et al., 2011); channels alternate between two possible means μ_1 and μ_2 with $\mu_1 > \mu_2$. Channels A and B are shown in blue and green, and channel C is shown in red; their signal means are μ_A, μ_B, and μ_C, respectively, with $\mu_A = \mu_B \neq \mu_C$

For these simulations, and following the protocol described in Tsetsos et al. (2011), an absorbing trial boundary was implemented. This boundary worked by forcing selection at the end of the trial period if no option reached the threshold before then. The rule for this selection is that the channel with more accumulated evidence - closest to reaching the threshold - be the one chosen.

We repeated the protocol presented previously using the MSPRT implementation by Bogacz and Gurney (2007). In their implementation, due to the link with the basal ganglia, a high threshold is not one that is large but one that is closer to zero since input evidence crosses this boundary from above. Evidence accumulation will start at 1.1 with three competing channels since at time 0 the Output value will be $ln(e^0 + e^0 + e^0)$.

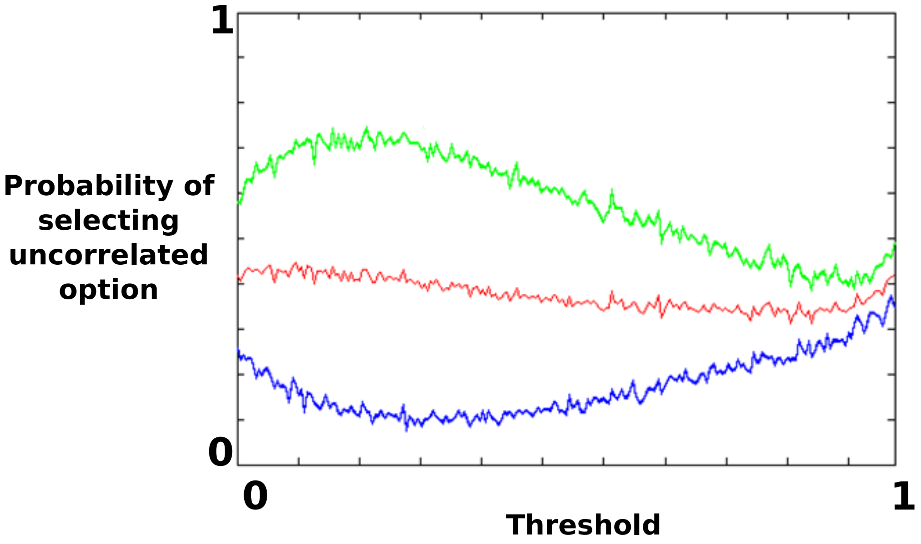


Figure 2.3: Probability of selecting uncorrelated option for a range of thresholds. Green line represents this probability for C when trial starts in phase 2, blue line in phase 1, and red line represents the average of green and blue.

Figure 2.3 shows the result for two sets of experiments, one when trial starts in Phase 1 and other when it starts in Phase 2. In this figure, the probability of selecting option C (uncorrelated) is plotted for each threshold, with the green line representing this probability when trial starts in phase 2, the blue line when trial starts in phase 1, and red line the average of green and blue.

The results show a clear primacy effect, which indicates that the likelihood of choosing the option supported in the beginning of the trial is higher.

In addition to the primacy effect, the results also show a preference for the uncorrelated option, C. This is consistent with the results obtained from the basal ganglia decision models in which a selection limiting effect is observed. This means that two closely matched options (similar means) will compete more strongly with each other and take more time to get selected than a single option would. These results contribute to our hypothesis that the MSPRT algorithm, as described by Bogacz and Gurney (2007) can withstand higher variability in inputs.

2.1.3 Robust decision making in non-stationary environments

In trial-based $MSPRT$ with a single decision, when the threshold is reached a decision is made, integration stops, and evidence integrators are reset to zero. This is reflected in the decision boundary being *absorbing* - no further processing takes place after a threshold crossing. In contrast, in a temporally extended, non-stationary situation, we require multiple decisions to be made with options which were once selected, being subsequently de-selected in favour of others. This demands that the decision boundary is now *transparent* - the trajectory described by an option may recross the boundary in the opposite direction from that which caused its selection. In addition, subsequent decisions are allowed for any option by threshold crossing for selection. An example of this kind of decision making is shown in fig. 2.4. Between a selection and subsequent deselection event (threshold crossing) of a channel, it is deemed to be in a 'selected state'.

Building on the previous example of a task where a participant has to decide which of a set of lights is the brightest, in this case, an example of a task with a transparent boundary is one in which the participant is presented with two separate lights. Let's assume that for 10 second light one is brighter, the participant will observe both lights for a period of time (accumulating evidence) and eventually decide that light one is the brightest (when the accumulated evidence for one option crosses the threshold). After the 10 second period, light one is dimmed, making light two the brightest; at this point, we expect the participant to revert their decision and select light two as the brightest; considering that the decision threshold is maintained, option one must have its evidence increased (assuming our MSPRT-like paradigm where selection results from threshold crossing from above) and recross the threshold (getting deselected), whilst light two decreases its evidence and crosses the threshold (getting selected). In order for the task presented before to occur, the threshold must be transparent so it allows subsequent selections and deselections; this does not constitute a change of mind as described by Resulaj et al. (2009) since it is associated with an evidence reversal (new decision) and find a parallel in the basal ganglia, which has the ability to switch from a selected channel to another following its increase in input evidence (Prescott et al., 2006). Henceforth, therefore, we assume a transparent decision boundary, unless otherwise stated.

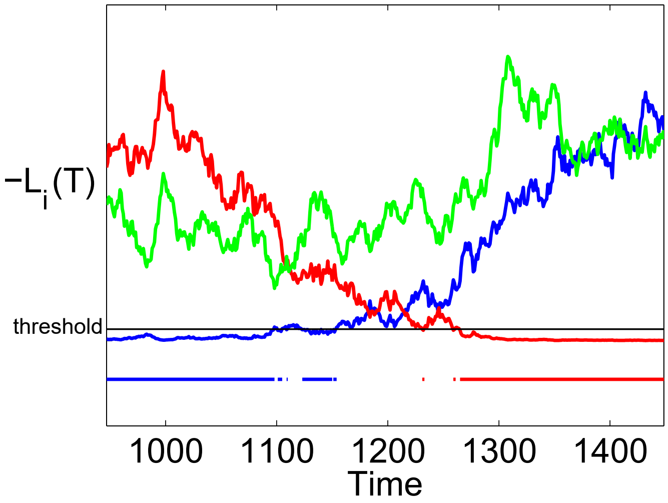


Figure 2.4: Decision making with a transparent boundary. There are three competing channels (shown red, green and blue) and their outputs, $\bar{L}_i(T), i = 1, 2, 3$; see equation (2.6). These signals cross the threshold (solid black line just above zero) from above and below, with threshold crossing from above resulting in a decision. Decision states are shown in the horizontal lines at the bottom, with the colour indicating the selected channel at any time. Note that not all times are associated with a decision. The signals are representative of those in a trial of fig. 2.7 for the fixed rectangular window (700 ms) in epoch 2 where $\Delta\mu = \mu^+ - \mu^- = 1$ and $\sigma = 0.3$. For this example the threshold was increased from 0.05 to 0.3 to elicit more selection and better illustrate the switch of selection from one option to another.

Interestingly, not all decision mechanisms are capable of using a transparent boundary. To do so requires that the expectation value of the dynamic variable which has to reach threshold (posterior probability for MSPRT) can have its direction reversed. This may occur via competitive processes, as are present in the drift diffusion model (Ratcliff and Rouder, 1998) , or a decay process, as in the leaky competing accumulator (LCA) (Usher and McClelland, 2001). However, the relevant variables in the race model (Vickers,

1970) are monotonically increasing and so this model does not allow de-selection with a transparent boundary.

To see that the MSPRT is a candidate for non-stationary decision making, suppose that channel j, having been selected at time T_j, has no further evidence accumulated after $T_0 > T_j$. Then for $T > T_0$, using (2.6), $\bar{L}_j(T) = -y_j(T_0) + Z(T)$. This is monotonic increasing in T since the baseline level for evidence accumulation is above the threshold, which means that once a channels has no more evidence accumulated, the outdated evidence will slowly be discarded bringing it back to baseline. The channel will therefore re-cross the threshold from below (be deselected) at some time $\bar{T}_j$; that is, channel j becomes deselected at $\bar{T}_j$.

Now consider channel i which is the next to be selected. The time to decision for i depends on the rate of increase of evidence $y_i(T)$ but also on the competition term $Z(T)$. Since Z(T) depends on the evidence accumulated for the competing channels, it now incorporates contributions from all previous 'experience' of the decision mechanism and therefore represents an ever increasing impediment to threshold crossing by i, as this experience grows. Even if $y_i(T)$ increases at a high rate, so does $Z(T)$ and if a competing channel has previously had a high rate of increased evidence, its contribution to the increase of $Z_i(T)$ will still be present after the evidence stops supporting it. It is as if the inability to 'forget' old information is causing subsequent decisions to be progressively harder.

One way of alleviating this problem is to force the 'forgetting' of previous evidence using a temporal *sampling window* and, conceptually, the simplest window form is rectangular. This is equivalent to defining the accumulated evidence as a convolution of the evidence samples and a rectangular function $w(t)$ with non-zero duration T_w. Thus, dropping the channel suffix for simplicity (with * indicating convolution),

$$y(T) = (x * w)[t] \tag{2.7}$$

$$= \sum_{t=1}^{\infty} x(t) \times w(T-t) \tag{2.8}$$

where

$$w(t) = \begin{cases} 0, & \text{if } t < 0, \\ g, & 0 \leq t \leq T_w \\ 0, & t > T_w \end{cases} \tag{2.9}$$

and g is the gain defined in (2.3). Using these limits on non-zero $w(t)$ we have

$$y(T) = \sum_{t=T-T_w}^{T} g \times x(t) \tag{2.10}$$

The decision mechanism now incorporates evidence from a fixed number of samples only over this window. Recall that convolution may be thought of as multiplying one function by another which 'slides' over the first; note that, if $T_w = T$, and $x(t)$ is stationary, this reduces to the normal MSPRT.

In order to test the effectiveness of the moving window method with non-stationary inputs, we compared the results of the original MSPRT (with no windowing) to variations with two fixed windows of short (300ms) and long (700ms) duration. We used three channels, with input means $\mu_1(t), \mu_2(t), \mu_3(t)$ defined by two hypothesis testing means

μ^+, μ^-, with $\mu^+ > \mu^-$. Two epochs were defined according to

$$\begin{array}{llll} \text{Epoch 1} & 0 \leq t < 1000: & \mu_1(t) = \mu^+, & \mu_2(t) = \mu_3(t) = \mu^- \\ \text{Epoch 2} & 1000 \leq t < 2000: & \mu_1(t) = \mu_3(t) = \mu^-, & \mu_2(t) = \mu^+ \end{array} \tag{2.11}$$

The standard deviation of the inputs, σ was fixed throughout for both channels at $\sigma = 0.33$. We explored performance under different difficulty of decision by varying $\Delta\mu = \mu^+ - \mu^-$ over the closed interval [0.15,1.5] sampling with an increment 0.15.

We also investigated the effect of the decision threshold θ, whose variation could show features not apparent in the case of stationary inputs. To fix the threshold range, we note that, at the start of the experiment, the evidence for each channel will be zero. Substituting $y_i = 0$ into (2.6) for three channels we have $\bar{L}_i(T_w) = \ln(\exp(0) + \exp(0) + \exp(0)) = \ln(3) \approx 1.1$; this is referred to as the *zero-evidence limit* and places an upper bound on theta as there should be no threshold greater than that associated with no evidence. For our threshold range we therefore set the maximum threshold at 0.5 guaranteeing accumulation of evidence before the threshold was reached. An order of magnitude range was used so that θ was varied over the closed interval [0.05 0.5] sampling at intervals of 0.05.

We evaluated performance by examining the outcome at the end of epoch 2, since this represented a typical decision in a non-stationary environment where the evidence may contain irrelevant contributions from a previous epoch. The results, averaged over 1000 experiments (each with a different instance of the stochastic input), and at each combination of $\Delta\mu$ and θ, are summarised in fig. 2.5. A 'correct outcome' was one in which the channel with the larger input was selected at the end of epoch 2; that is, its negative log-posterior, $\bar{L}_i$, was less than θ. The other possible outcomes were: 'incorrect selection' (the channel with the smaller input was selected); 'no selection' (neither channel had $\bar{L}_i < \theta$); and 'multiple selection' (more than one channel had $\bar{L}_i < \theta$).

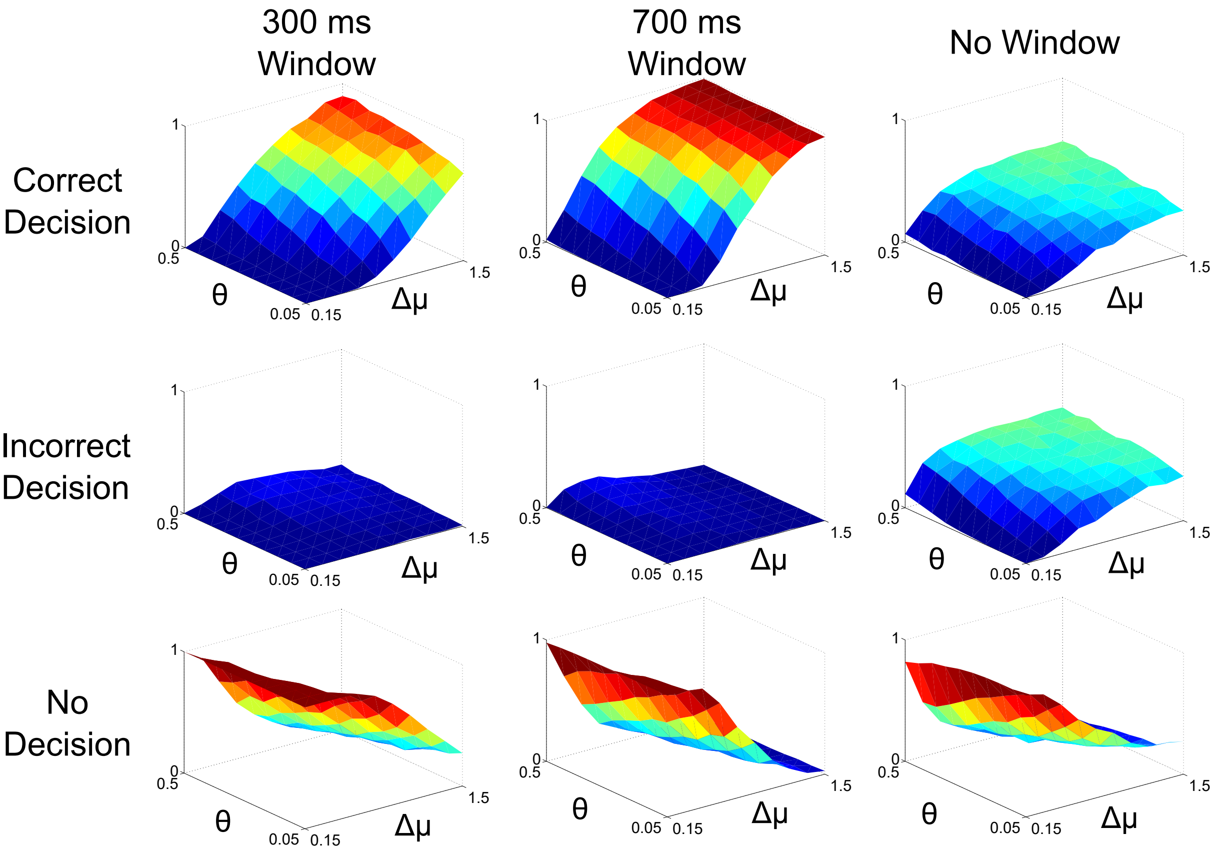


Figure 2.5: Performance of windowed and non-windowed MSPRT as a function of decision difficulty (defined by $\Delta\mu = \mu^{+} - \mu^{-}$; see text) and threshold θ. The input protocol is that in equation 2.11.First row shows the probability P, (fraction of 1000 experiments) of performing a *correct* decision at the end of the trial for each of the cases: 300ms window, 700ms, and no window. Second row shows the corresponding probabilities of performing an *incorrect* decision. Third row shows the probabilities for reaching no decision at the end of the trial.

The top row of panels in fig. 2.5 shows the probability P, of obtaining the correct outcome (fraction of the 1000 runs). The non-windowed (original) MSPRT performs at roughly chance levels for most of the parameter range. This is because the decision mechanism is being called on to work in epoch 2, but is 'contaminated' with erroneous evidence from the first epoch. In contrast, for both small and large window sizes, there is a threshold range which allows better performance than the non-windowed MSPRT.

In particular the 700 ms window shows an almost perfect performance for a significant range of problem difficulty and threshold. Thus the ability to 'forget' prior evidence can be advantageous.

Further, the 300ms window benefits from somewhat higher thresholds for low problem difficulty (large $\Delta\mu$). Since the increase in threshold is moderate, it does not result in a significant increase in incorrect selection. The benefit of larger thresholds may seem counter-intuitive because, for the un-windowed MSPRT, smaller thresholds require more evidence to reach, and are associated with lower error-rates. However, with the windowed variants, if the threshold is too low, the time required to reach it may exceed the window time, and somewhat higher thresholds can force selection, albeit not always correct. In addition, reporting the correct decision at the *end* of epoch 2, is forgiving of errors earlier in the epoch (threshold crossing with the wrong outcome) which might be present with larger thresholds. The results therefore reflect a tension between these conflicting features.

Further insights into the decision process with non-stationary inputs are supplied by the second and third rows of panels in fig. 2.5 which show (for each of the window options) the probabilities of making the wrong selection, and no selection respectively. Except at very small $\Delta\mu$ and threshold, the un-windowed MSPRT makes an incorrect decision almost at chance for most of the parameter space. In contrast, the errors of performance for the short (300ms) and long (700ms) windows MSPRT are mainly a failure to reach a decision (fig. 2.5, row 3, columns 1 and 2), which is not surprising given its limited time to do so.

There is some benefit for the 700ms window from large thresholds with very hard problems (small $\Delta\mu$). However, in the main, for the larger window size, the performance is fairly independent of threshold (up to values close to the zero-evidence bound). Thus, across both small and large windows, moderately large thresholds are preferred. Further increasing the threshold will not result in increase reward since this will increase both correct and incorrect selection.

For all MSPRT variants, the tendency for no-selection (third row in fig. 2.5) is most pronounced for hard problems and small thresholds because both conditions demand longer accumulation times. This is true even for the non-windowed case which may require longer than 2000ms (the trial time).

In sum, windowing appears to confer an advantage for the MSPRT in non-stationary environments by allowing prior, irrelevant information to be discarded. For windows of smaller duration, the counter-intuitive benefits (with the current performance criterion) of somewhat larger thresholds (but below the zero-evidence limit) are explained by the limited amount of evidence available with small windows - such thresholds avoid the impasse of no selection. For larger windows (comparable to the characteristic time-scale of the stimulus) performance is fairly insensitive to the threshold but, where there is dependence, larger values (but removed form the zero-evidence limit) are preferred for the same reasons.

2.2 Problem-dependent windowing improves performance

Consider the performance of the windowed MSPRTs in fig. 2.5. Both the long (700ms) and short (300ms) duration windows show significantly better performance than those obtained using the original MSPRT, although the longer window performs better, overall, than its shorter counterpart. Thus, conceiving of the original MSPRT as using an effectively 'infinite' window, these results suggest a modal progression of behaviour with window size in which there is a preferred size for best performance. Further, the preferred size is related to the epoch size (here 1000ms). Thus longer/shorter epochs would give rise to longer/shorter preferred window size. However, selecting a window size based on the epoch duration is not a practical proposition for an autonomous agent who has no knowledge of the timing of the input transitions.

An alternative approach to selecting window size is indicated by examining the performance 'landscape' over the problem space. Thus, for 'easy' problems (with large $\Delta\mu$), the shorter (300ms) window performs almost as well as the longer (700ms). For harder problems (with small $\Delta\mu$), there is significant benefit in using windows longer than 300ms.

We might therefore ask the question: what is the *adequate* window duration, for each degree of problem difficulty (here, determined by $\Delta\mu$) that does not result in large drop in performance? It is more realistic to suppose that an agent could estimate problem difficulty by, perhaps, separately sampling the inputs over a short, fixed window to estimate their means (we have, indeed, investigated this possibility and it does appear viable).

Based on the observations above, we assume that the adequate window size scales with the decision time for a stationary problem with three inputs. To investigate this, we therefore used a model with an un-windowed MSPRT with three, constant-mean input channels. The two 'losing' channels had mean μ^- and the winning channel μ^+, as defined previously; $\Delta\mu$ was varied in the interval $[0.6, 2.6]$.

For simplicity of the model, the standard deviation of the noise (σ) was kept constant at 0.33, however, it is worth noting that this parameter also increases and decreases problem difficulty depending on whether its value is high or low.

Although we did not conduct an exhaustive search, good results in the non-stationary experiments were subsequently obtained for a 15% error rate with the un-windowed MSPRT. For consistency, we also used the threshold for this error rate (0.05) in the subsequent, non-stationary input experiments. Further, using a 5% error rate in the power law determination resulted in only a 1% decrease in mean reward (compared with that for 15% rate).

The results, averaged over 1000 runs at each value of $\Delta\mu$, are shown in left panel of

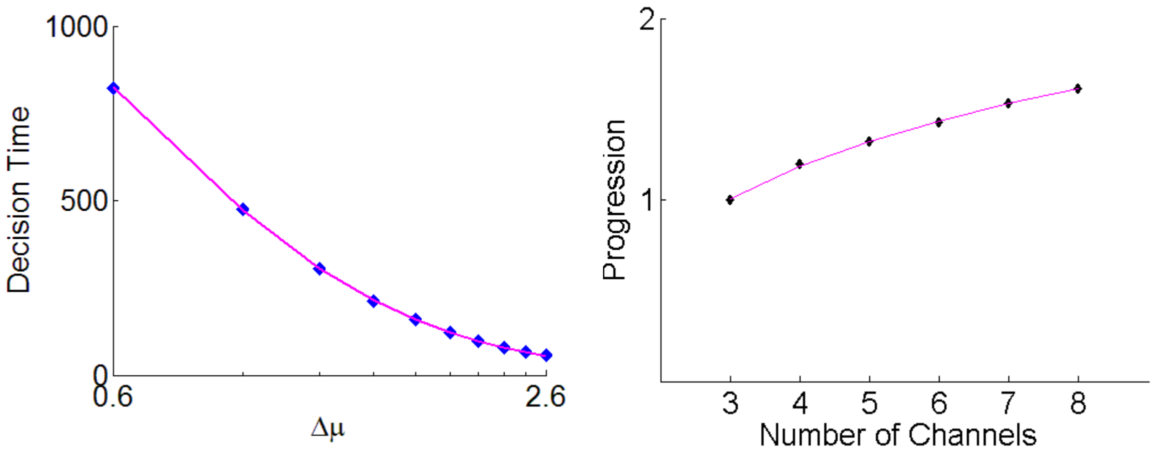


Figure 2.6: Left panel depicts decision time versus decision difficulty. Blue dots represent mean decision time against $\Delta\mu$ (log axis) in a three choice task with stationary inputs over 1000 runs. Curve shows power law fitted to decision time. Right panel depicts the proportional increase in decision time for a number of competing channels (N), with reference to $N = 3$.

fig. 2.6. The data were well fitted by a power law relating the mean decision time and $\Delta\mu$

$$DT(\Delta\mu) = 445.7\Delta\mu^{-1.98} \tag{2.12}$$

We then used this to determine the problem-dependent window size $T_w(\Delta\mu)$ by simply putting $T_w(\Delta\mu) = DT(\Delta\mu)$. This method of determining $T_w(\Delta\mu)$ is robust. In other experiments with $N > 3$ (not reported here), we used the $N = 3$ relation in (2.12) with good selection performance. Moreover, we have shown that the relations like (2.12), specifically derived with $N > 3$, are approximated quite well by simply scaling that for $N = 3$ in (2.12).

The relationship between decision time for $N = 3$ and $N > 3$, depicted in the right panel of fig. 2.6 was fitted by a logarithmic curve, and therefore the decision time for any number of channels and $\Delta\mu$ can be found using equation 2.13:

$$DT(\Delta\mu, N) = [0.6175 * \ln(N) + 0.3289] * 445.72 * \Delta\mu^{-1.976} \tag{2.13}$$

The value of window size generated using the particular curve for each N and that of equation 2.13 had an average difference of 0.75%, with this value never exceeding 2%. This equation is a useful tool since it doesn't require computation and storage of N equations, which can become extremely memory and time consuming as N increases.

Thus far, our evaluation of performance has rested solely on the outcome at the end of an epoch. However, in order to better gauge performance in terms of the decision *throughout* an epoch, we introduce a metric whose value is governed by the *duration* of a correct decision. Thus, at each time step i within an epoch, we set a Boolean variable r_i to 1 if the correct channel (and only this channel) had been selected, and -1 if an incorrect channel was selected. The metric R_j was then defined as sum of all such values over the epoch j.

$$R_j = \sum_{i \in J} r_i \tag{2.14}$$

where J is the set of times steps in epoch j. This captures the notion that early (and sustained) correct decisions are better than late, or incorrect ones (with reference to equation (2.11), the correct outcome in epochs 1 and 2 are selection of channels 1 and 2 respectively).

Experiments with the problem-dependent window were conducted over a range of differences in mean $\Delta\mu = \mu^+ - \mu^-$, and standard deviation σ, common to both the inputs. We therefore refer to $R_j(\sigma, \Delta\mu)$ with respect to a particular parameter pair. It is also useful to refer to the mean of R_j over a set of experiments. Thus, if L, M are the experimental sets of σ, and $\Delta\mu$ respectively, we define the mean reward in epoch j

$$\overline{R_j} = \langle R_J(\sigma, \Delta\mu) \rangle_{L,M} \tag{2.15}$$

As suggested by the notation, the metric in (2.14) might be interpreted as some kind of cumulative 'reward' for correct decision making – early, correct decisions lead to larger values of accumulated reward – and we will therefore refer to it in this way. Note that the symmetry of R about zero penalises performance at chance which will yield expected reward values of zero.

In the following experiments the threshold was fixed at 0.05, the value used to obtain Equation (2.12). The parameter sets L, M (defining variations in $\sigma, \Delta\mu$) comprised the 10, equi-spaced values in the closed intervals [0.1,0.55], [0.15,1.5] respectively, and each pair of parameters was used in 1000 trials. Experiments were done with a fixed window of 700ms, no window, and the variable window defined by (2.12); the results are shown in fig. 2.7.

For epoch 1 (first column of fig. 2.7), there is little difference in performance (reward) across window variation. However, the results in the first epoch are not typical of the non-stationary decision process since, during this time, there is no prior evidence to contaminate the decision (and since both the fixed and un-windowed mechanisms accumulate more evidence in epoch 1, they are able to outperform the variable window mechanism).

In contrast, for epoch 2, the reward with fixed or problem-dependent (adjustable) window is significantly higher than that with no window. Unlike epoch 1, in epoch 2, a prior decision has to be reversed and prior evidence is an impediment to making the new, correct choice, thereby favouring the shorter problem dependent windows. The problem-dependent and fixed (700ms) windows produce similar mean rewards of 146, and 156 respectively. However, the performance of the problem-dependent window is better over large parts of the 'easier' problem space.

In order to further understand the behaviour with the problem-dependent window, fig. 2.8 shows a dissection of the behaviour in epoch 2 for the case shown in fig. 2.7 (row 3, column 2) into each decision possibility at the end of the epoch. Thus, in the 'correct

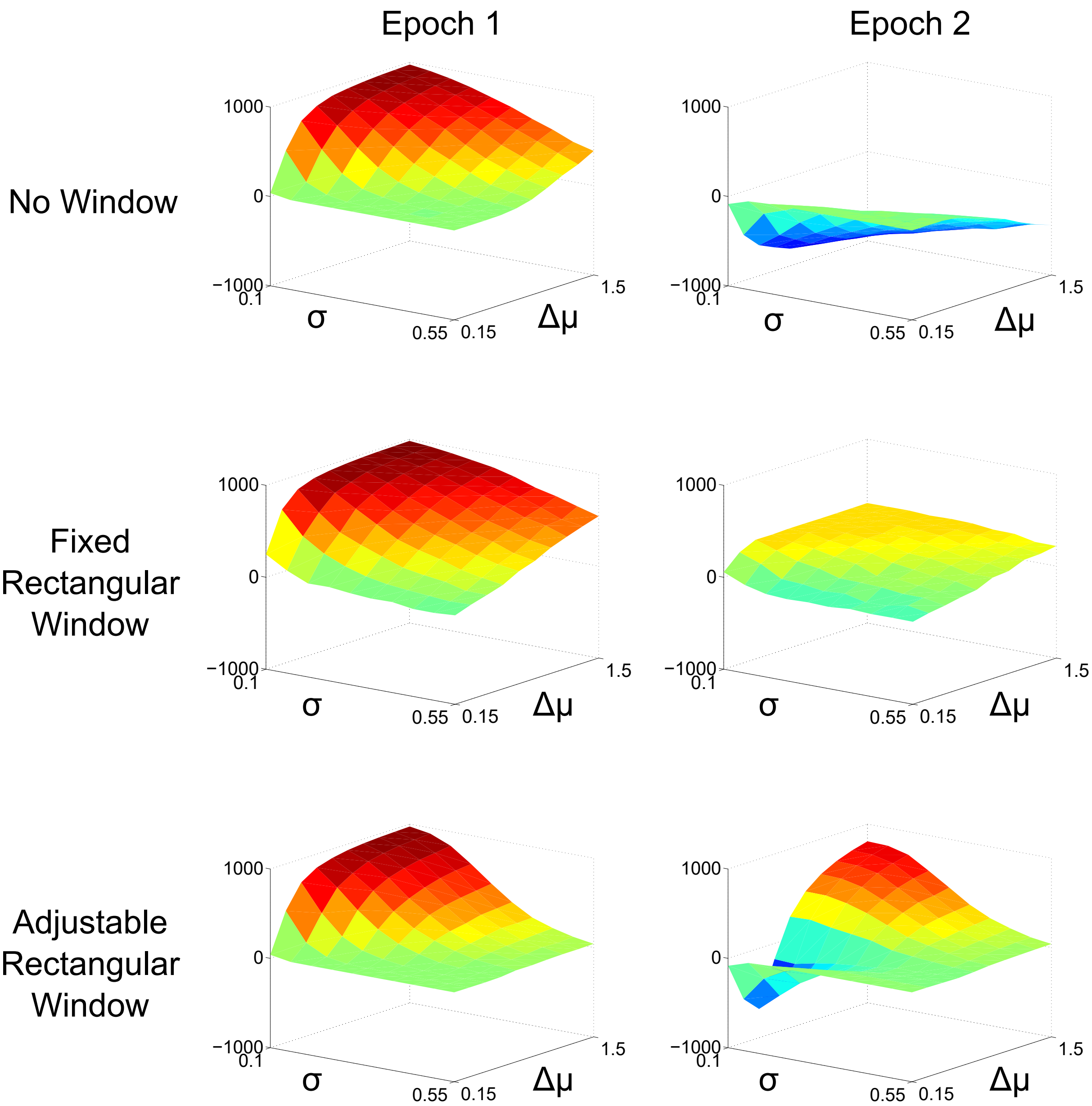


Figure 2.7: Cumulative reward in the experiment of equation 2.11; average of 1000 runs. First and second column for R_1, R_2 (epochs 1, 2) respectively. Each triplet of panels shows the case for: no window, fixed (700ms) window, and problem dependent window (determined by (2.12))

decision', channel 1 is selected at the end of epoch 1, in an 'incorrect decision' it is channel 2 or 3, and 'no selection' implies no channel has crossed threshold. Multiple selection was omitted since it did not occur in significant numbers. It is clear that the failure to reach a reward maximum is largely due to failure to make a selection of any kind.

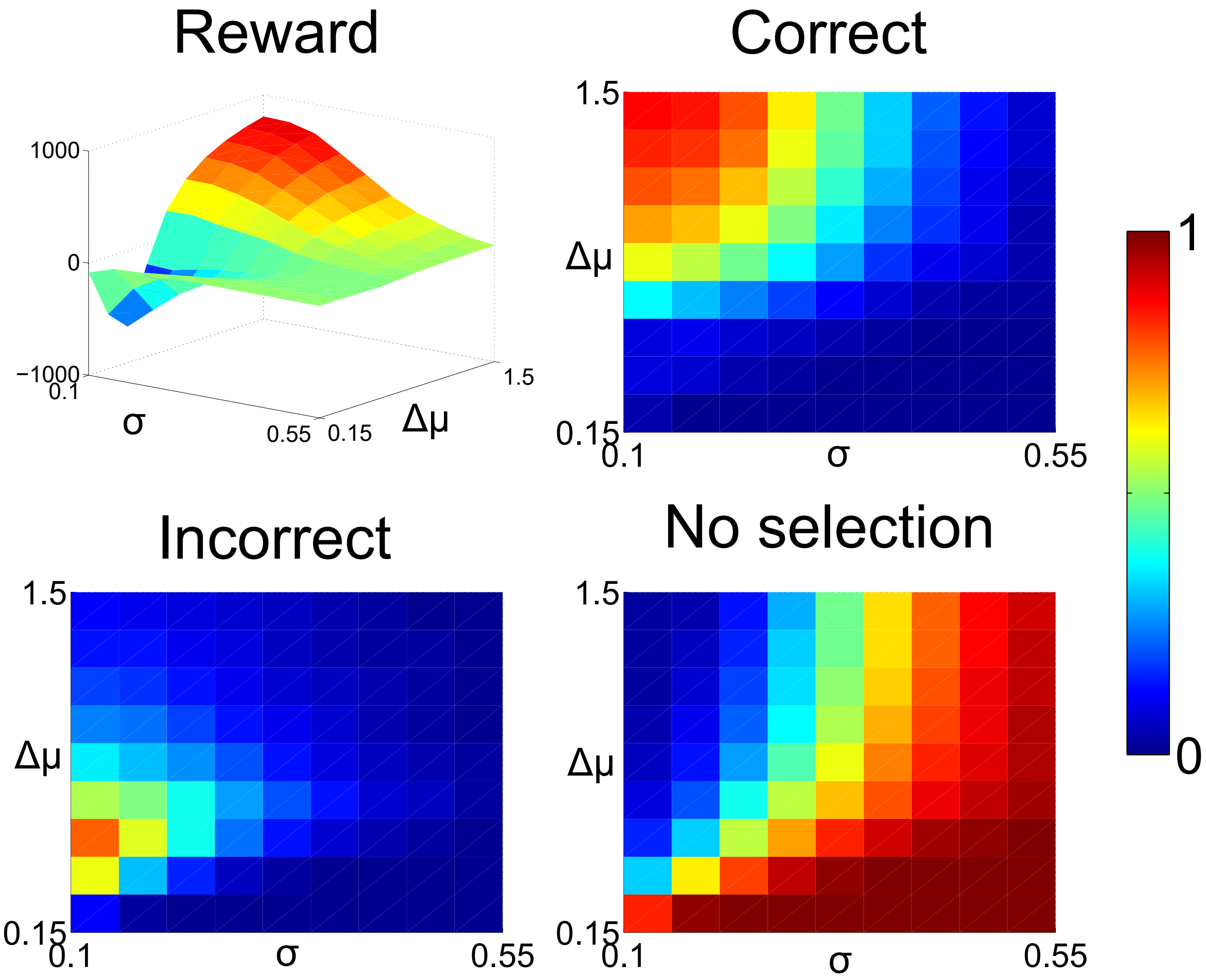


Figure 2.8: Cumulative Reward at end of epoch 2 for an adjustable rectangular window, from fig. 2.7 (row 3, column 2). Row 1 column 2 depicts the probability of correct selection, row 2 column 1 depicts the probability of incorrect selection, and row 2, column 2 depicts the probability of performing no selection. Range of values shown on the right pertains to decision figures (not to reward figure)

The results presented previously correspond to averages over multiple sets of parameters, which are not particularly informative about what was taking place in particular locations of the parameter space. To address this, we broke down the very high level averaged figures presented previously and generated single trial and multiple trial outputs for particular pairings of parameters. These pairings represented the more extreme situations which can sometimes be masked by averaging over a large parameter space; and that normally contain relevant information about the functioning of the model.

In this case and because our varying parameters were σ and $\Delta\mu$, the pairings were situations where: $\Delta\mu = 0.5, \sigma = 0.1$ (Low $\Delta\mu$, Low σ); $\Delta\mu = 1.5, \sigma = 0.1$ (High $\Delta\mu$, Low σ); $\Delta\mu = 0.5, \sigma = 0.4$ (Low $\Delta\mu$, High σ); and $\Delta\mu = 1.5, \sigma = 0.4$ (High $\Delta\mu$, High σ).

Two separate analysis were performed, one looking at the output of the algorithm and other at the selection output. The Decision Output was defined in the following way: if a channel ch is selected at time t, $DO(ch, t, sim) = 1$, if not, $DO(ch, t, sim) = 0$. Both quantities were measured and presented for a particular trial and over 100 trials (average value at time t).

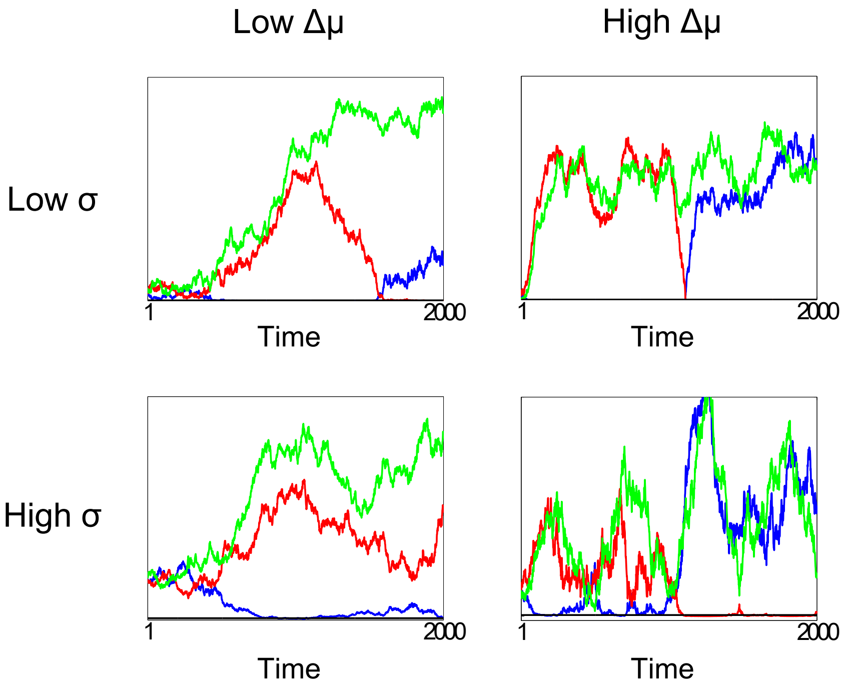


Figure 2.9: Figure shows the Output over time of each competing channel (red, green and blue) on a single trial. Line in black represents the decision threshold and is more visible in trials where σ is high, due to the overall amount of evidence.

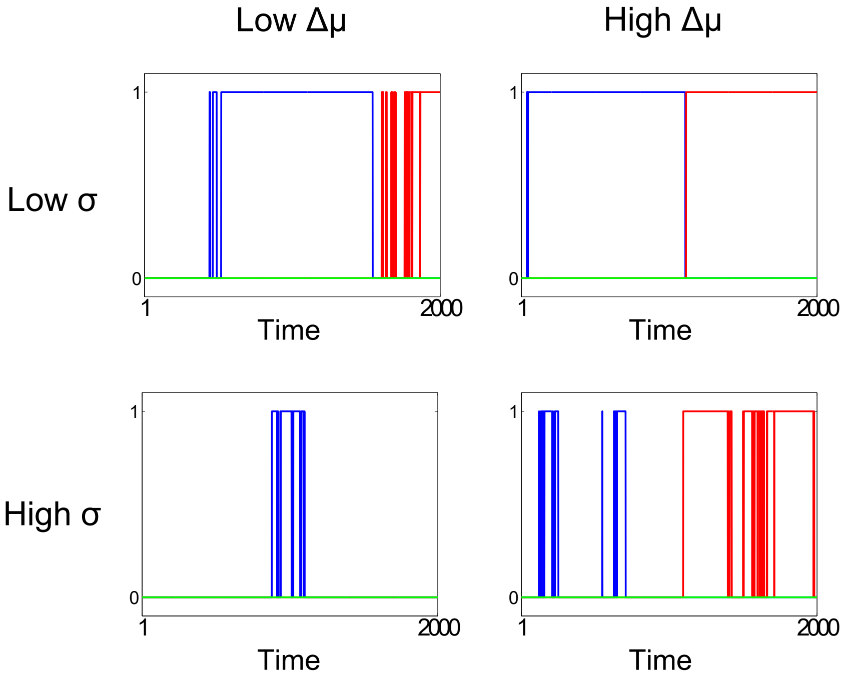


Figure 2.10: Figure shows the Decision Output over time of each competing channel (red, green and blue) on a single trial. A value of one for a particular channel at time t indicates that this channel was the one selected at that time..

Figures 2.9 and 2.10 show the Output and the Decision Output for the various pairings of parameters for a single trial. In order to comprehend these results, one must remember that the selection of an option in this algorithm is done when a signal crosses the threshold from above. This Decision Output was shown in fig. 2.4 where the solid line at the bottom represented the chosen channel at a particular point in time. The colour of the line indicated which channel was selected. The blue and red lines represent the selection output for the blue and red channels, respectively. If the output of the algorithm goes down and crosses the threshold, its decision output will go up (getting selected)

At first we can note that the algorithm shows robustness to variations in $\Delta\mu$. This is the case since it still maintains high decision output levels (albeit the wrong decision for a small portion of the trial), even when $\Delta\mu$ is low (more difficult decision). This can be observed in the first column of figs. 2.9 and 2.10.

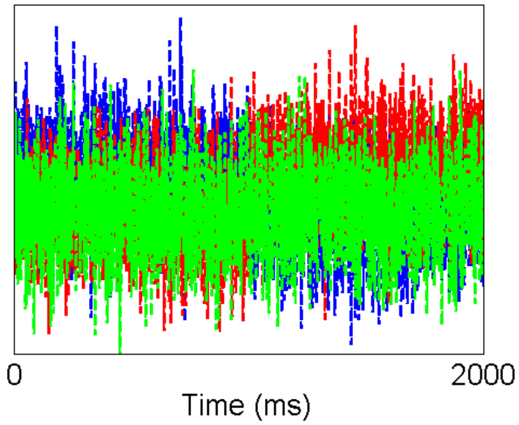


Figure 2.11: Example of input evidence where σ and $\Delta\mu$ are very closely matched, resulting in a large intersection between the evidence for channels 1 and 2.

The algorithm is, however, more sensitive to variations in noise, with the limit case (high σ) showing the poorest performance. Nonetheless, one should note that in the limit situation of high noise, the σ is almost of the same magnitude as the lowest $\Delta\mu$ which in turn results in a very large area of intersection of the probability density functions of the competing signals. This results in input evidence like that in fig. 2.11.

Figures 2.12 and 2.13 show the Output and Decision Output for the same sets of parameters as figs. 2.9 and 2.10, but in this case an average of 100 trials.

These results confirm the earlier observations in figs. 2.9 and 2.10, where the algorithm showed sensitivity to high levels of noise. One can note that in extreme situations such as those shown in column 2 and row 2 of figures figs. 2.12 and 2.13, the decision output for High σ and high $\Delta\mu$ drops to around 0.3 which means that the correct channel was selected only around 30% of the time; this does not constitute chance level or indicate wrong selection since of another channel the selection output for the other channel is 0. In this case, the algorithm simply did not select any option - error rate was maintained and amount of selection was decreased overall.

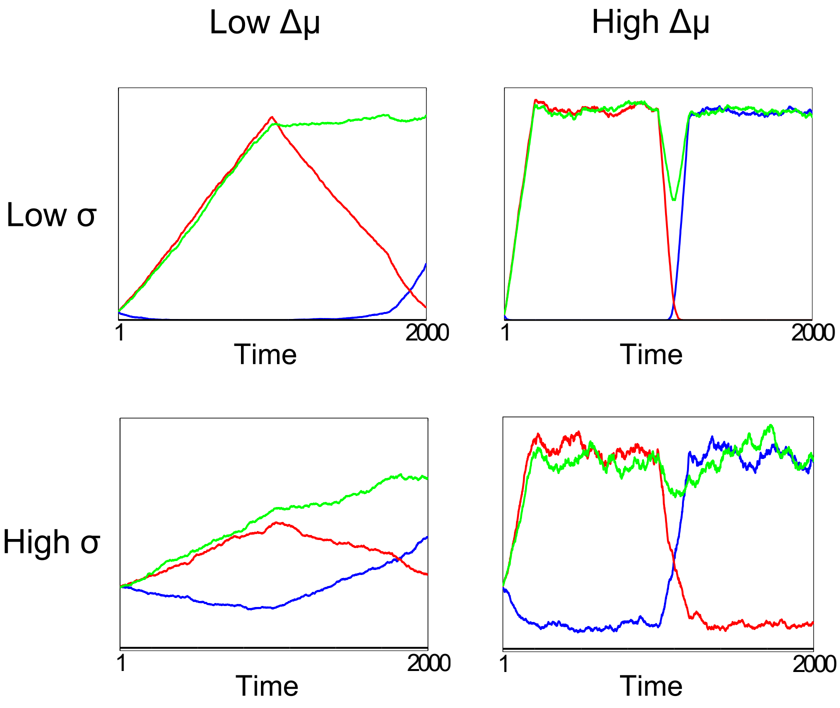


Figure 2.12: Figure shows the average Output over time for each competing channel (red, green and blue) over 100 trials. Line in black represents the decision threshold and is more visible in some figures but not others due to the overall amount of evidence.

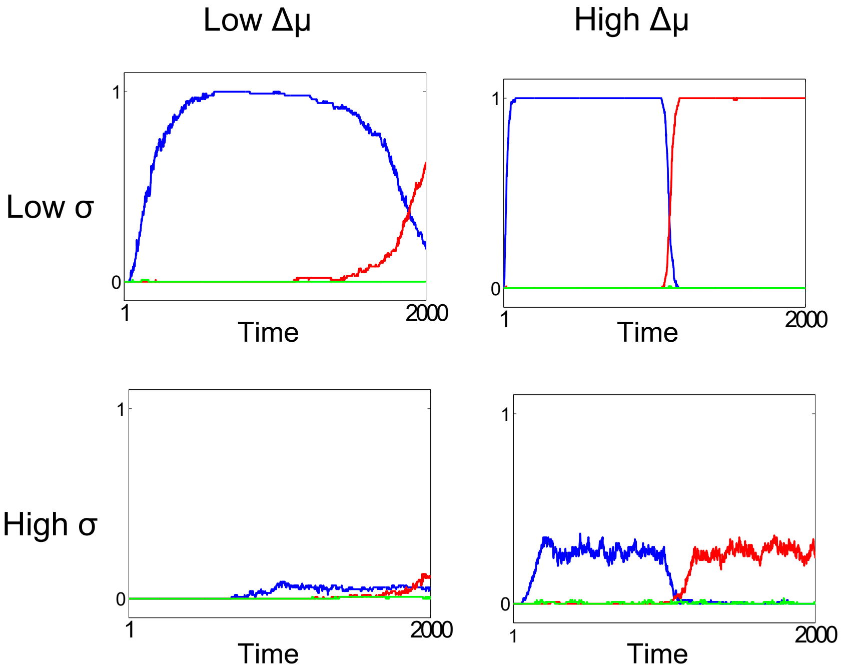


Figure 2.13: Figure shows the average Decision Output over time of each competing channel (red, green and blue) over 100 trials. A value close to 1 at time t indicates that the particular channel was selected on most trials at that point in time.

2.2.1 Robustness under moderate change in gain

Thus far the accumulated evidence has made use of a windowed sum of data samples $x(t)$, multiplied by a gain g, where $g = (\mu^+ - \mu^-)/\sigma^2$. The evaluation of this gain would seem to require the mechanism has knowledge of its own inputs (means and standard deviation). However, in Bogacz and Gurney (2007) it was shown that the performance of the basic MSPRT mechanism was independent of the gain, as long as it was larger than the nominal value given here. We therefore sought to find the effect of changing the gain on performance in the windowed mechanism described here. This was done by taking the gain g defined above and multiplying by a factor k, with $k = 0.5, 5, 10, 20$.

The results are shown in fig. 2.14. In particular, fig. 2.14 shows that, with $k = 0.5$ (when the gain is below nominal), the performance is compromised. With $k = 5$, there is an improvement in performance (average reward) for large $\Delta\mu$ and σ. For $k > 5$ there is no benefit in performance, but for small σ, there is a severe reduction in reward. This trend is also more prevalent for small $\Delta\mu$. This phenomenon can be traced to a numerical issue caused by having to take the exponential of the gain-multiplied accumulated evidence in (2.5). If circumstances conspire to make this evidence large, these exponentials can be too big for reliable representation in the machine. In particular, for small σ, the nominal gain g takes its largest values over the problem space and, multiplication by relatively large k increases this tendency further. In addition, for small $\Delta\mu$, the adaptive sampling window is longer, and the accumulated evidence can become relatively large.

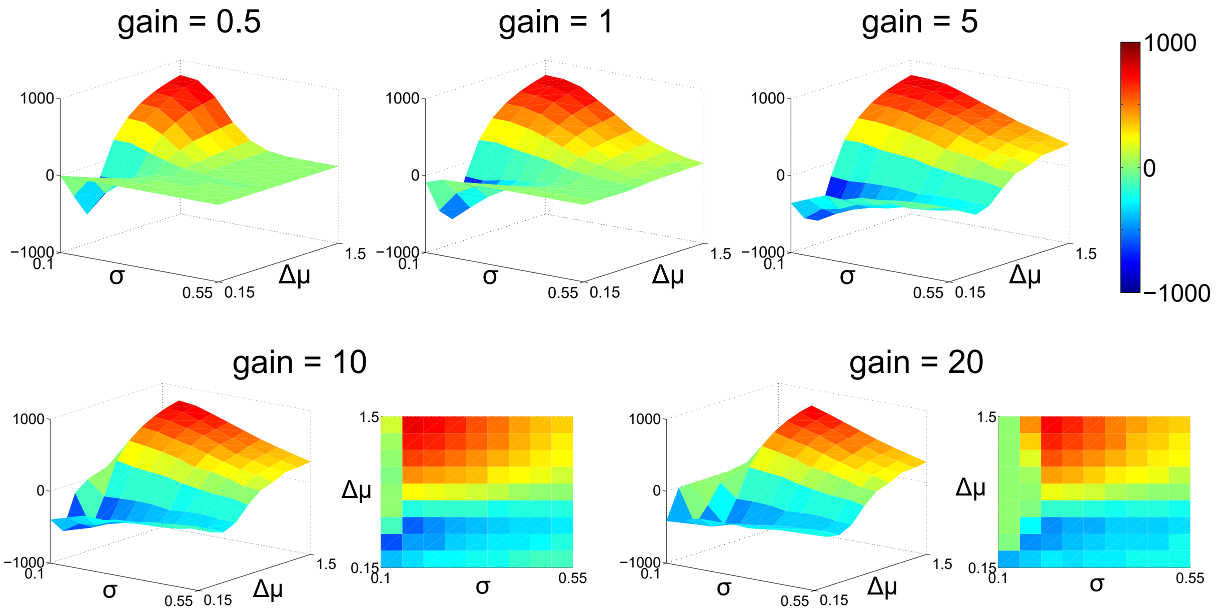


Figure 2.14: Effect of gain on performance with the adaptive window. Cumulative reward R in epoch 2 using the protocol in Equation 2.11 shows the original result with gain $g = (\mu^+ - \mu^-)/\sigma^2$. Row 1, columns 1 and 3 are for the case when this gain is multiplied by 0.5 and 5 respectively. Row 2, columns 1,2 and 3,4 are for the case when the gain is multiplied by 10 and 20, respectively, both in 3D form and as a 2D contour plot, to highlight features otherwise masked by the view in 3D (which is consistent with the other panels)

For areas of the graph where $\Delta\mu$ is maximum, the decision is interpreted as easy even though in some cases σ is high. Since window size is only dependent on $\Delta\mu$, the difficulty of the decision is underestimated and the window is set to a a small length, which leads to decreased reward from non-selection. In this case, a small increase in gain can have a positive effect on correct decision without visible effects on incorrect decision since it compensates for the small window size by increasing the overall level of evidence. A slightly higher gain can also compensate for sub-optimal values of threshold.

Notwithstanding this, simply ensuring the gain is between its nominal value g, and up to around five times this value, does not cause any deterioration of performance.

2.2.2 Exponential window geometry

Here we address the issue of the memory-intensive requirements of a rectangular sliding window noted above. Consider a non-rectangular window defined by the exponential function

$$w(t) = \begin{cases} g\exp(-\lambda t), & \text{if } t \geq 0 \\ 0, & \text{if } t < 0 \end{cases} \tag{2.16}$$

where $\lambda = 1/\tau_w$ and τ_w is a characteristic time constant measured in time steps, and g is given in (2.3). Then convolving this with the data $x(t)$ (channel suffix suppressed), and using the limit on non-zero $w(t)$, we obtain the equivalent of (2.10)

$$y(T) = g\sum_{t=1}^{T} x(t)\exp[-\lambda(T-t)] \tag{2.17}$$

At the next time step

$$\begin{aligned} y(T+1) &= g\sum_{t=1}^{T+1} x(t)\exp[-\lambda(T+1-t)] \\ &= \exp(-\lambda)\left\{ g\sum_{t=1}^{T} x(t)\exp[-\lambda(T-t)] + gx(T+1)\exp(\lambda)\right\} \\ &= \exp(-\lambda)y(T) + gx(T+1) \end{aligned} \tag{2.18}$$

Assuming many time steps for the exponential to decay significantly, $\tau_w \gg 1$ or $\lambda \ll 1$, so that $\exp(-\lambda) \approx 1 - \lambda$. Using this in (2.18)

$$y(T+1) = (1-\lambda)y(T) + gx(T+1) \tag{2.19}$$

Thus, there is a very simple, memory efficient update rule for accumulating evidence with the exponential window: reduce or ‘decay’ the current accumulated evidence $y(T)$ by a factor $1 - \lambda$ and add in the new (gain weighted) evidence sample.

To fix τ_w (and hence λ) in a way which enabled meaningful comparison with previous results, we chose to use a problem-specific adaptive window and obtain λ using the power law in (2.12). Thus we set $\lambda = 1/DT(\Delta\mu)$. Figure 2.15 shows the resulting reward using the protocol in Equation 2.11. For ease of comparison, corresponding results for the adaptive rectangular window are show alongside.

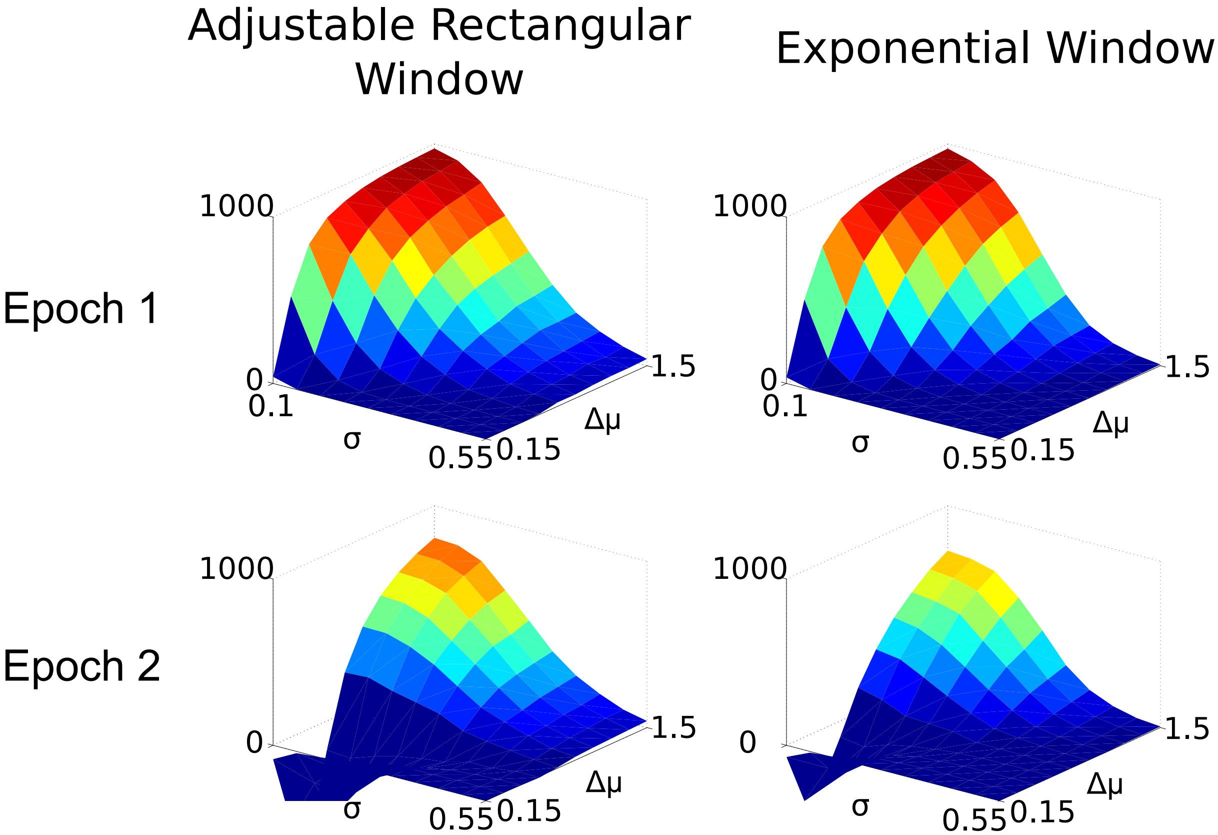


Figure 2.15: Comparison of results from problem specific windows with rectangular and exponential form. Cumulative reward R using the protocol in Equation 2.11. Column 1, rows 1,2 show average reward of 1000 runs with rectangular window, for epochs 1 and 2. Column 2, rows 1,2 show average reward of 1000 runs with exponential window, for epochs 1 and 2.

In order to test whether setting $\lambda = 1/DT(\Delta\mu)$ yielded optimal results, we introduced a scaling term n between 0.25 and 5 whereby:

$$\lambda = n \times \frac{1}{DT(\Delta\mu)} \tag{2.20}$$

The quality of results from each term (n) was measured by averaging the overall reward for all sets of parameters (σ and $\Delta\mu$). Due to the different dynamics of each epoch, different terms resulted in higher or lower levels of mean Reward, $\overline{R_j}$, as defined in Equation 2.15, according to the epoch analysed.

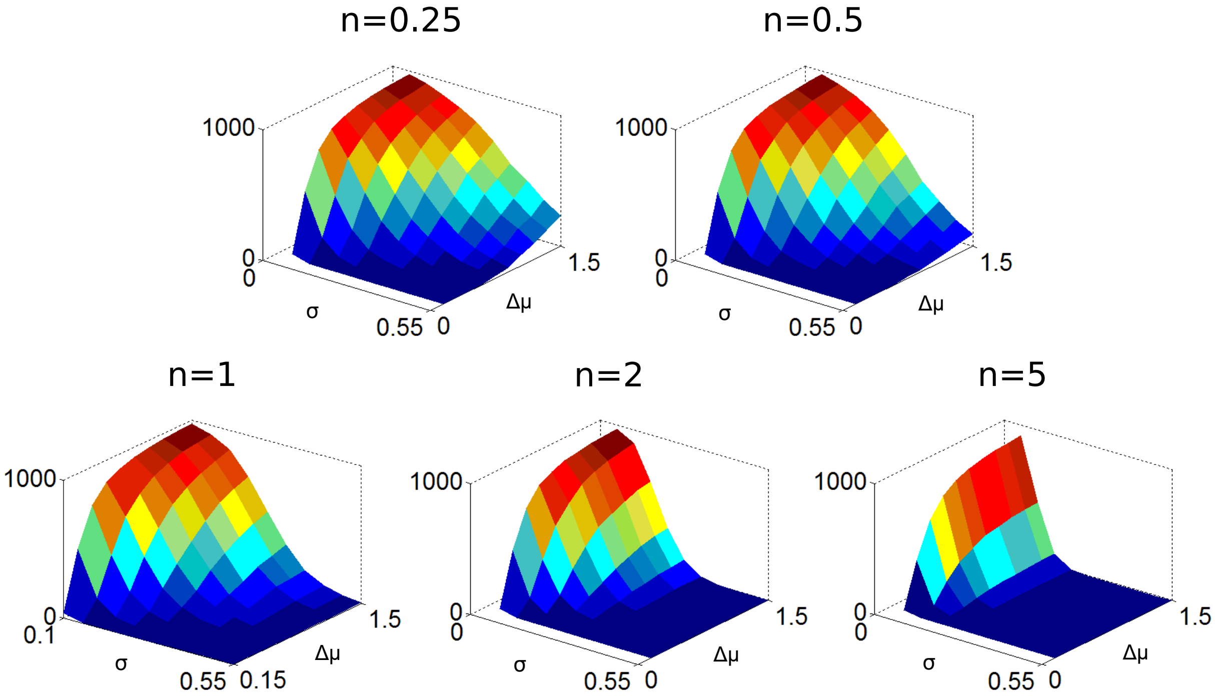


Figure 2.16: Cumulative Reward at end of epoch 1 with an exponential multiplier (λ) given by Equation 2.20 for a range of σ and $\Delta\mu$.

Results for epoch 1 can be observed in fig. 2.16 where each panel relates to a value of n for a range of σ and $\Delta\mu$ parameters.

For this epoch, smaller terms performed better, with $\overline{R_1} = 351$ for n=0.25, $\overline{R_1} = 320$ for n=0.5 and $\overline{R_1} = 266$ for n=1 versus $\overline{R_1} = 197$ for n=2 and $\overline{R_1} = 106$ for n=5.

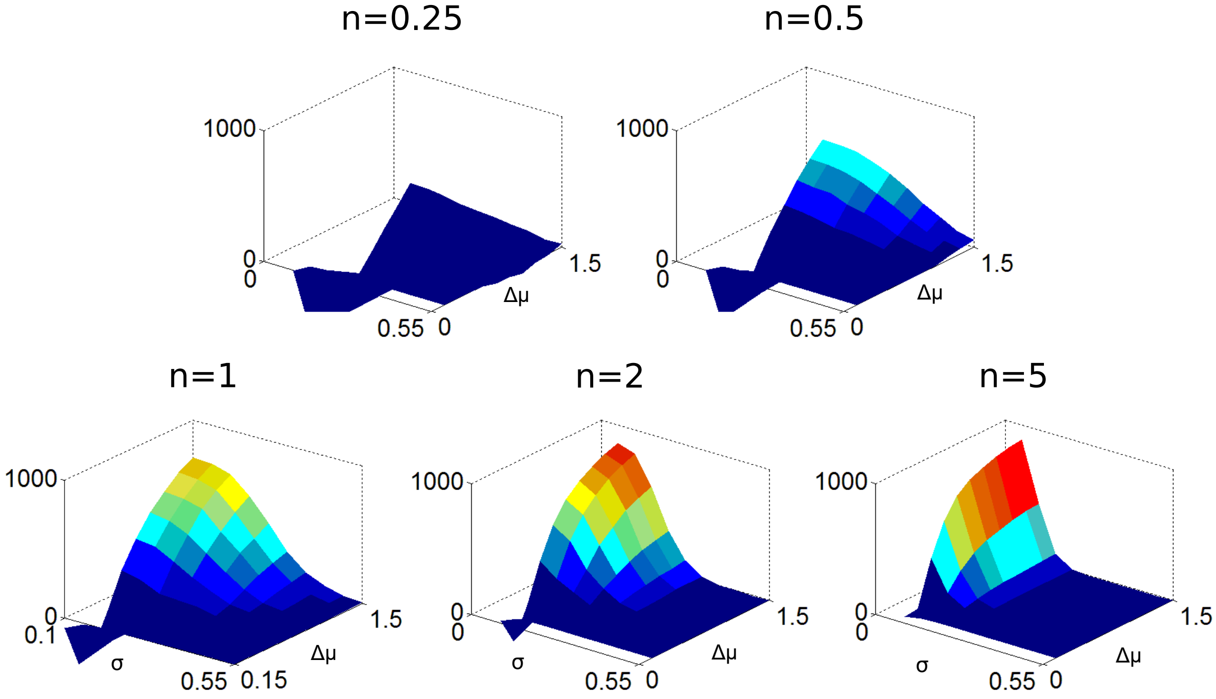


Figure 2.17: Cumulative Reward at end of epoch 2 with an exponential multiplier (λ) given by Equation 2.20 for a range of σ and $\Delta\mu$.

Results for epoch 2 can be observed in fig. 2.17 where similarly to fig. 2.16 each panel relates to a value of n for a range of σ and $\Delta\mu$ parameters.

Unlike in epoch 1, in epoch 2, larger terms performed better, with $\overline{R_2} = -102$ for n=0.25, $\overline{R_2} = 27$ for n=0.5, $\overline{R_2} = 110$ for n=1, $\overline{R_2} = 128$ for n=2 and $\overline{R_2} = 89$ for n=5.

The results show that larger n terms tend to perform better for later epochs, and smaller n terms for earlier ones. Since we are interested in more difficult decisions with previously accumulated evidence, larger terms are more interesting. For this epoch, n=2 yielded the best results.

The fact that this results in the best performance for the later epoch is in concordance with the idea that an adaptive window is more robust to changes in difficulty of decision, especially when such window is adapted to this.

During the review process of a paper containing part of the research included in this chapter (Nunes and Gurney, 2016), our attention was drawn to work of which were not aware (Veliz-Cuba et al., 2015). In this publication, the authors implemented a non linear stochastic model of evidence accumulation which used exponential discounting as a means of discarding outdated evidence. In their implementation, however, the evidence discount rate is dependent on the frequency of environmental changes and assumes that those are always known to the decision maker.

We believe this to be fundamentally different from our own implementation of the exponential window, with the only common point being the geometry of the discounting window. In our implementation, λ is calculated through the use of a power law which depicts decision time for a given number of competing channels and for a particular $\Delta\mu$, which we discovered empirically.

Chapter 3

MSPRT algorithm with dynamic input patterns

3.1 Can MSPRT-like models be used if inputs don't follow the theoretically assumed pattern?

In the previous chapter, we covered the scenario where the means were assumed to have the theoretical pattern described in Bogacz and Gurney (2007), where all losing channels had the same mean input; for this, only the winning channel and $\Delta\mu$ were allowed to vary. This was the first step in approximating a realistic input scenario. There we showed that the MSPRT algorithm could withstand the added variability with the addition of a windowing mechanism capable of discarding outdated evidence.

The next step towards ethological plausibility would be to assume that the competing channels do not follow the winner-loser pattern and so different input channels can have different means, but what does this mean for algorithms with MSPRT-like computations?

In terms of MSPRT computations and its parameters, a variation in signal means affects only the gain parameter, since it is based on the difference between the means of the winning (μ^+) and the losing channels (μ^-). However, since the derivation of the gain parameter was based on the assumption that the inputs to the decision mechanism follow a winner loser pattern with only two means; with this no longer being true, we can no longer assume that its structure is maintained. Lepora and Gurney (2012) explored the effect of multiple means outside the winner-loser pattern and described a general Bayesian algorithm which worked in identical manner to the MSPRT model adopted by us throughout this thesis, when evidence is encoded as general log likelihoods.

In order to maintain the simplicity of our model without adapting it to work in a more generic way such as that described in Lepora and Gurney (2012), and by varying the signal means in a way that is outside the formalism established to derive the gain, we accept the loss of formal rigour and in turn set out an empirical study to analyse the effect of having different means on the overall speed and accuracy of the mechanism.

From a decision making stand point, the smaller the difference between competing options, the more difficult the decision is considered, and so, in theory, given the smallest difference between the winning channel and any of the losing ones, we would have found the most challenging scenario for the algorithm.

The scenario where all losing channels have the same input mean is the hardest for the decision mechanism since as soon as some of the options start to decrease their input mean, the decision is made easier because the competition is lessened. By setting the $\Delta\mu$ parameter to this value, we can adapt the mechanism to the 'worst possible scenario' and therefore better comprehend its weaknesses.

By controlling for the worst possible scenario, we are being conservative in terms of the error rate, since a small $\Delta\mu$ leads to a small gain and so all options take longer to reach the decision threshold. However, in the previous chapter we showed that is it actually advantageous to increase the gain slightly in order to increase the number of decisions, even if at the expense of the error rate. We also showed that an increase in magnitude of gain up to 10x yielded the highest reward. Consequently, even if we are calculating the gain based on estimated measures of the signal means, there is still a large margin for error.

There are two ways in which the gain can be wrongfully estimated, depending on whether we under or overestimate the value of the signal means.

If the estimated gain is higher than it should be, the competing options will reach the threshold more rapidly. This, in turn, will result in an increased number of errors, but also potentiates increased selection. This increased selection turned out to be advantageous since even though it boosted the error rate, the overall amount of correct selection was also boosted.

Conversely, if the gain is under-estimated, its value will be suboptimal, which will decrease the amount of evidence accumulation, therefore resulting in the algorithm taking longer than necessary to reach the threshold. This increase in decision time will decrease the overall amount of selection (both correct and incorrect) with no negative impact on error rate.

In the following section we will examine exactly what the effect of changing the gain (while maintaining the winner-loser input scenario) is on the decision time and compare the results for the MSPRT as described in Bogacz and Gurney (2007) with two different alternative implementations where extra variability is added to the inputs.

3.2 Decision times maintained for non-ideal inputs

In their 2007 paper, Bogacz and Gurney (2007) showed that the MSPRT model was robust to changes in gain since it did not deviate significantly from its asymptotical optimality when the gain was varied from the optimal value $g*$. For their implementation, there was no increase in decision time for $g > g*$ and even though there was an increase for $g < g*$, the decision time was still lower than that of the Usher and McClelland (2001) model. Notwithstanding the fact that the inputs to our model don't always follow the winner-loser pattern theorised, the gain parameter was redefined as the smallest difference between the competing channel means, as explained in previous sections.

The progression of decision time versus $g/g*$ can be seen in figure 3.1 for both models, with the solid line representing the Bogacz and Gurney (2007) and the dashed line the Usher and McClelland (2001) model.

For this part of our work, and to assess to what degree we were deviating from the asymptotic optimality of MSPRT, we used the same methods as Bogacz and Gurney (2007) with our implementation of the MSPRT.

For each instance of $g/g*$ in a range from 0.06 to 3 we found, numerically, 10 thresholds which resulted in a $1\% \pm 0.2\%$ error rate. These thresholds were then used as a parameters in the model and the decision time recorded and averaged to construct the decision time data points in the figure. These results were obtained with N=10 competing channels.

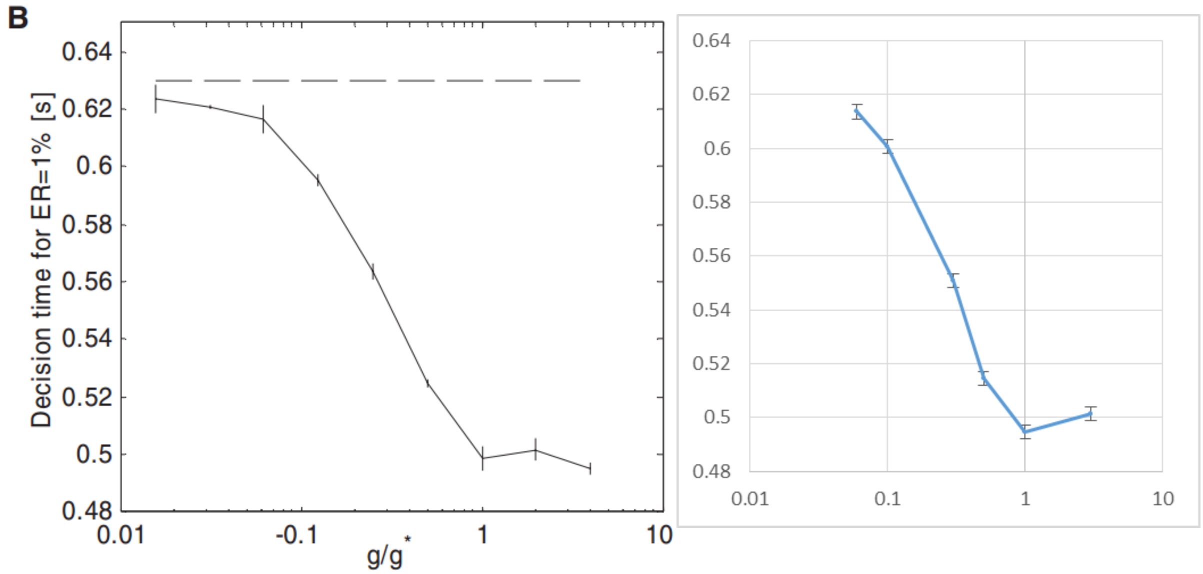


Figure 3.1: Robustness to change in gain. Left panel shows decision time versus fraction of optimal gain for the Bogacz and Gurney (2007) model, right panel shows results for our MSPRT-like model.

The results from our model are shown on the right panel of figure 3.1, and can be directly compared to those from the Bogacz and Gurney (2007) model, on the left panel of the same figure.

One can note that in both implementations the evolution of decision time for $g/g*$ follows the same pattern, in that it decreases as the gain increases with only a slight increase for $g/g* > 1$. Another aspect common to both is that the decision time is always lower than that of the Usher and McClelland (2001) model. The decision times for our model are in fact consistently lower than those of the original Bogacz and Gurney (2007) model.

The previous set of results demonstrate that even with added variability - where the means were allowed to be time-varying, while maintaining the pattern of 1 winner and $N-1$ equal losers - the decision times were not affected, displaying the robustness of our extended MSPRT model to changes in gain, like the original MSPRT.

The next step towards ethological plausibility would be to allow the input means to not only be time-varying but also for their patterns to change. In this case, for $N-1$ losing channels ,

$$\mu^2 \neq \mu^3 \neq ... \neq \mu^{k-1} \neq \mu^k, k = 2:N$$

For the initial simulation we used only two differences in means of the signals, so that

$$\mu^1 = \mu_{min} + 2 * \Delta\mu; \mu^2 = \mu_{min} + \Delta\mu; \mu^3 = ... = \mu^{k-1} = \mu^k = \mu_{min}, k = 2:N$$

The results from this implementation can be found in figure 3.2. As can be observed, the overall pattern of evolution of the decision time for g/g* was maintained, with higher gain resulting in lower decision time up to g/g*=1. However, when compared to both panels in figure 3.1, each individual decision time was lower consistently lower.

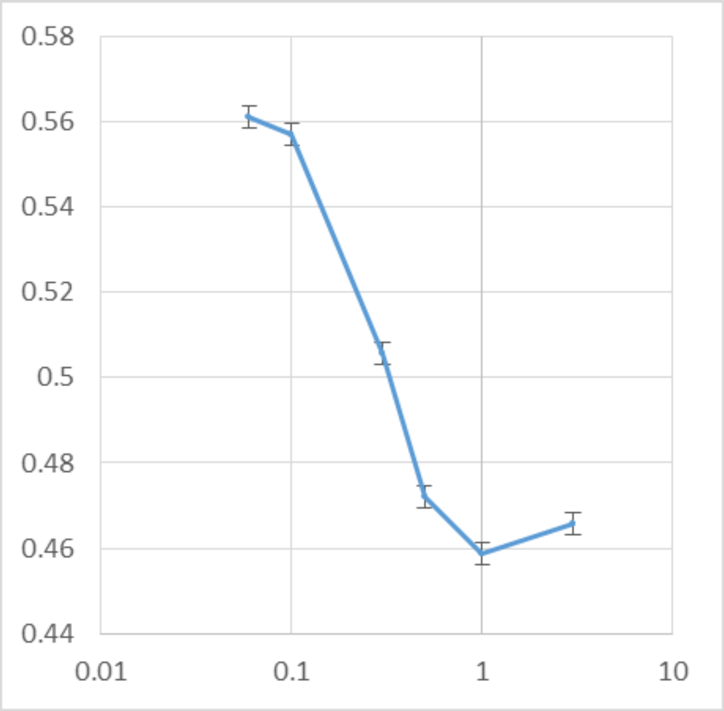


Figure 3.2: Robustness to change in gain with added variability

One might wonder how changing the input pattern to be different to the theorised can result in increased performance, but that is directly connected to the level of competition. As previously stated, when all the losing options are at the same level and taking into consideration that we already know the smallest $\Delta\mu$, we are already accounting for the most difficult scenario. In this case, as soon as one or more losing channels start to vary their means, i.e. decreasing them, the decision at hand becomes easier since the signals are more distinguishable than expected - the competition is decreased - and the winning option can reach the threshold more quickly.

3.3 Robustness under varying patterns of input

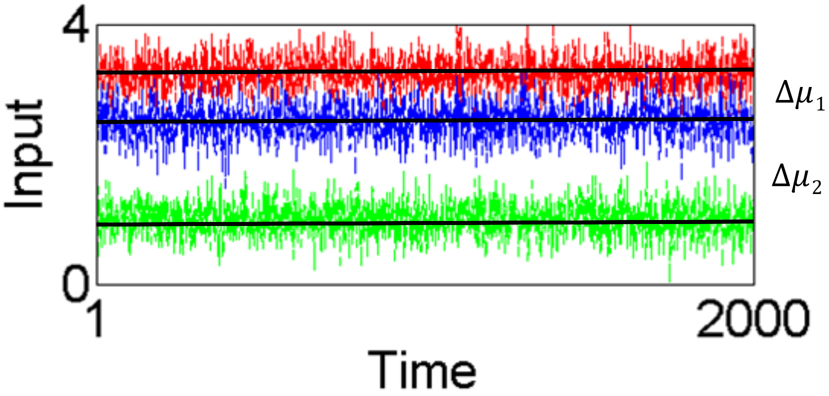


Figure 3.3: Input pattern with added variability, where $\mu_1 \neq \mu_2 \neq \mu_3$.

Consider the input pattern in figure 3.3, where the winning channel is represented in red, and the losing channels are represented in green and blue. If $\Delta\mu_1$ is the difference between the means of channels 1 and 2, with 2 being the winning channel, and $\Delta\mu_2$ is the difference between channels 2 and 3, with those being the two losing channels (in a 3 channel implementation). Assuming $\Delta\mu_1 > 0$ (there is a winning channel and therefore a decision to be made), if $\Delta\mu_2 > 0$, inputs follow a different pattern to the theorised and have the structure presented before

The more we increase $\Delta\mu_2$, the easier the problem becomes, since the competition is decreased, as shown before.

Table 3.1 shows the effect on decision time from the variation of the $\Delta\mu_1$ and $\Delta\mu_2$ parameters.

		$\Delta\mu_2$					
		0	0.3	0.6	0.9	1.2	1.5
$\Delta\mu_1$	0	-	-	-	-	-	-
	0.3	2501	2411	2182	2125	1746	1980
	0.6	1092	837	819	741	732	750
	0.9	349	320	298	279	255	255
	1.2	300	263	225	208	185	164
	1.5	180	178	133	132	128	126

Table 3.1: Average decision time in milliseconds (5% error rate) for a combination of $\Delta\mu_1$ and $\Delta\mu_2$

Through analysis of table Table 3.1, one can note that when $\Delta\mu_2$ is kept constant, the decision time consistently decreases with the increase of $\Delta\mu_1$. This result was expected since an increase in $\Delta\mu_1$ effectively translates in more pronounced isolation of the winning channel against the losing ones. This increased isolation will in turn lower the difficulty of decision, and consequently the decision time.

Similarly, fixing $\Delta\mu_1$ and following the $\Delta\mu_2$ columns from left to right will result in decreased decision time as $\Delta\mu_2$ grows. This increase in $\Delta\mu_2$ distances the second losing channel from the first and the winning channels, and like the increase in $\Delta\mu_1$, reduces competition and therefore difficulty, translating into a lower decision time.

		$\Delta\mu_2$					
		0	0.3	0.6	0.9	1.2	1.5
$\Delta\mu_1$	0	-	-	-	-	-	-
	0.3	-	-4	-13	-15	-30	-21
	0.6	-	-23	-25	-32	-33	-31
	0.9	-	-8	-15	-20	-27	-27
	1.2	-	-12	-25	-31	-38	-45
	1.5	-	-1	-26	-27	-29	-30

Table 3.2: Percentage Variation in average decision time(5% error rate) from $\Delta\mu_2 = 0$ for each combination of $\Delta\mu_1$ and $\Delta\mu_2$

Table 3.2 shows the difference in percentage between the decision time for $\Delta\mu_2 = 0$ for each set of $\Delta\mu_1$ and $\Delta\mu_2 > 0$. Once again, these results show a clear tendency for decreased decision time for a fixed $\Delta\mu_1$ as $\Delta\mu_2$ increases.

		$\Delta\mu_2$					
		0	0.3	0.6	0.9	1.2	1.5
$\Delta\mu_1$	0	-	-	-	-	-	-
	0.3	-	-	-	-	-	-
	0.6	-56	-65	-62	-65	-58	-62
	0.9	-86	-87	-86	-87	-85	-87
	1.2	-88	-89	-90	-90	-89	-92
	1.5	-93	-93	-94	-94	-93	-94

Table 3.3: Percentage variation in average decision time(5% error rate) from $\Delta\mu_1 = 0.3$ for each combination of $\Delta\mu_1$ and $\Delta\mu_2 > 0.3$

Similarly, Table 3.3 shows the difference in percentage between the decision time for $\Delta\mu_1 = 0.3$ for each set of $\Delta\mu_2$ and $\Delta\mu_1 > 0.3$. These results support the conclusions taken before since once again we can observe that for the majority of parameter sets, the decision time decreases as we increase $\Delta\mu_1$.

Both results showed a decrease in decision time through the variation of the input patterns where $\Delta\mu$ was increased; this supports the previously posed hypothesis that the problem difficulty is decreased as the input patterns are changed from the constrained versions theorised.

3.3.1 Unconstrained inputs and window sizes

Despite the fact that the previous results stem from a less constrained MSPRT-like algorithm since the inputs do not follow the same pattern as theorised, there is no variability within the trial, i.e. inputs are constant over the period of time analysed. In order to combine the previous results with those in chapter two, we devised another experiment in which the inputs varied over time, with different patterns that weren't always the winner-loser pattern. Simulations were run using no window, a window of fixed length equal to trial size and an adjustable window depending on $\Delta\mu$ as defined previously.

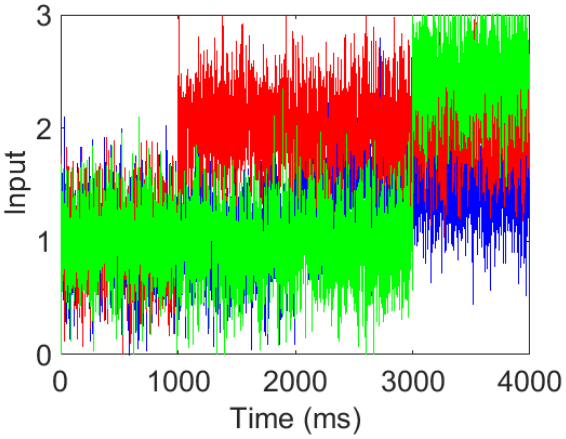


Figure 3.4: Input evidence with three competing options - 4 epochs

For this analysis we defined three epochs with four different desired outcomes:

- **Epoch 1 (0000-1000 ms):** No Selection
- **Epoch 2 (1000-2000 ms):** Option 1 (red) must be selected

- **Epoch 3 (2000-3000 ms):** Option 1 (red) must be selected; note that blue option has increased mean
- **Epoch 4 (3000-4000 ms):** Option 3 (green) must be selected

Results were displayed in surface plots of average reward (as defined previously) for each set of parameters (1000 trials) for each epoch. We will not present results from the 1^{st} epoch since in this period there are no winning options; σ was varied in a range of [0.05,0.55] and the $\Delta\mu$ in a range of [0.15,1.5].

By observing fig. 3.5, the results seem to contradict the assumption that the $MSPRT$ with the window always outperforms the unwindowed one; this is indeed the case since on the 2^{nd} Epoch the unwindowed $MSPRT$ gives a higher average reward than the algorithm with the dynamic window.

In the input pattern utilised, the 1^{st} epoch has no inputs apart from noise from the system so for the first 1000 ms the model accumulates only noise; since the unwindowed version of the algorithm never throws away information, all the accumulated noise is the difference between a weak signal reaching the threshold or not in the 2^{nd} Epoch.

The reason why the results in fig. 3.5 are not significant is because we are interested in running our algorithm continuously. The type of decision we are faced with, in this case, will only happen once in every trial, regardless of how long the algorithm runs for. This is because it is characterised as being a first decision with only noise preceding it. Every decision after this one will have some previous information about other decisions and not just background noise, which means that the conditions will differ from these. Even if later on in the trial all the inputs were to be reduced to μmin again, this would no longer help the unwindowed algorithm because inputs would have been building up and a significantly larger amount of evidence would be necessary to contradict an existing decision.

Unlike fig. 3.5, fig. 3.6 represent the typical conditions encountered when running the algorithm. In this case, a previous decision has been made and so it looks at the algorithm's ability to discard outdated information.

It is clearly visible that the windowed version of the $MSPRT$ outperforms the unwindowed one. At the same time, fig. 3.7 shows results similar to those in fig. 3.6, which indicates that the algorithm maintains its decision capabilities.

In the 4th epoch, the mechanism with a fixed window of 1000 ms decayed its reward levels, however always outperforming the unwindowed mechanism. This indicates, again, that the $MSPRT$ with some type of windowing mechanism always outperforms its unwindowed counterpart.

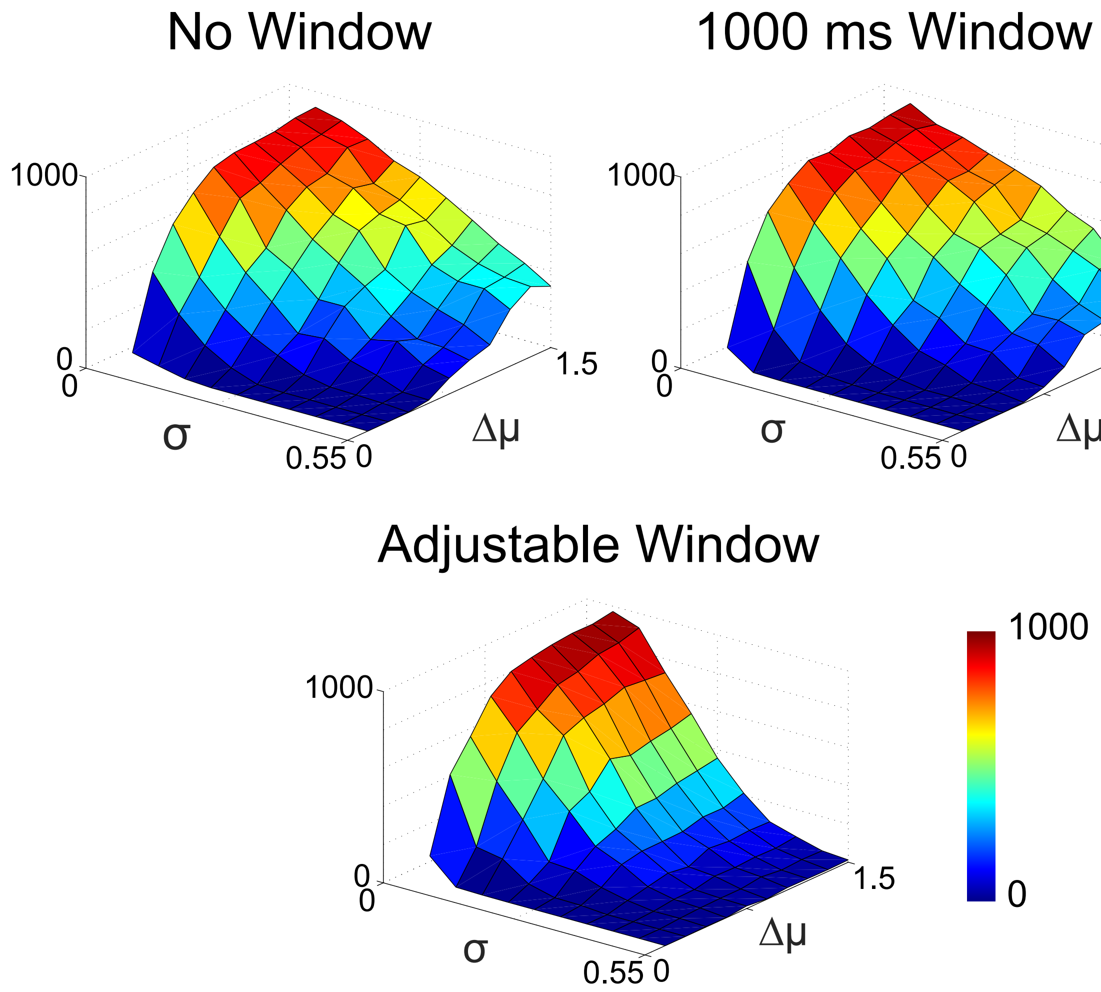


Figure 3.5: Average cumulative reward the second epoch (1000-2000ms) of fig. 3.4; average of 1000 runs. Top left, top right and bottom correspond to algorithm with no window, fixed (1000ms) window, and adjustable window, respectively

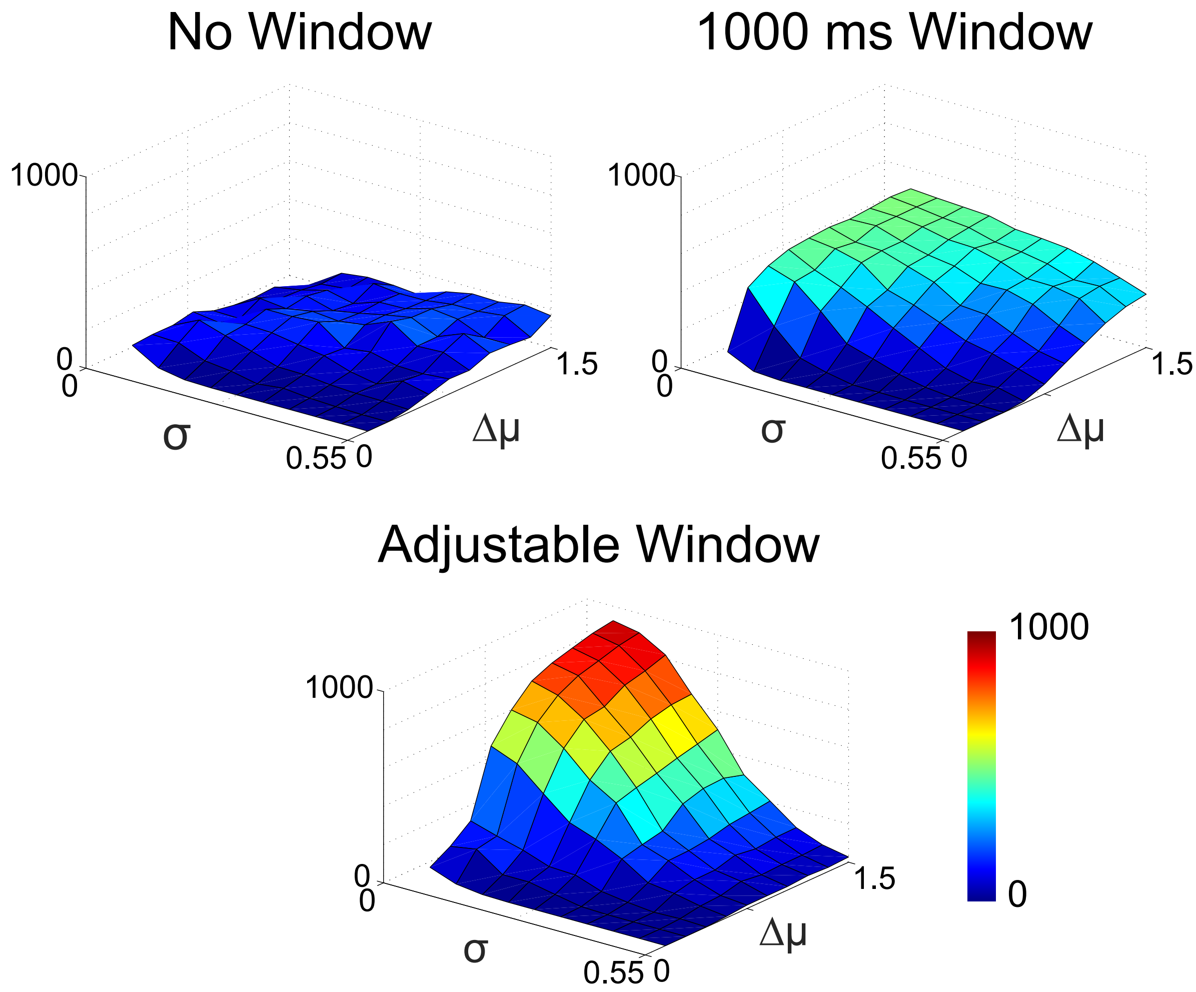


Figure 3.6: Average cumulative reward the third epoch (2000-3000ms) of fig. 3.4; average of 1000 runs. Top left, top right and bottom correspond to algorithm with no window, fixed (1000ms) window, and adjustable window, respectively

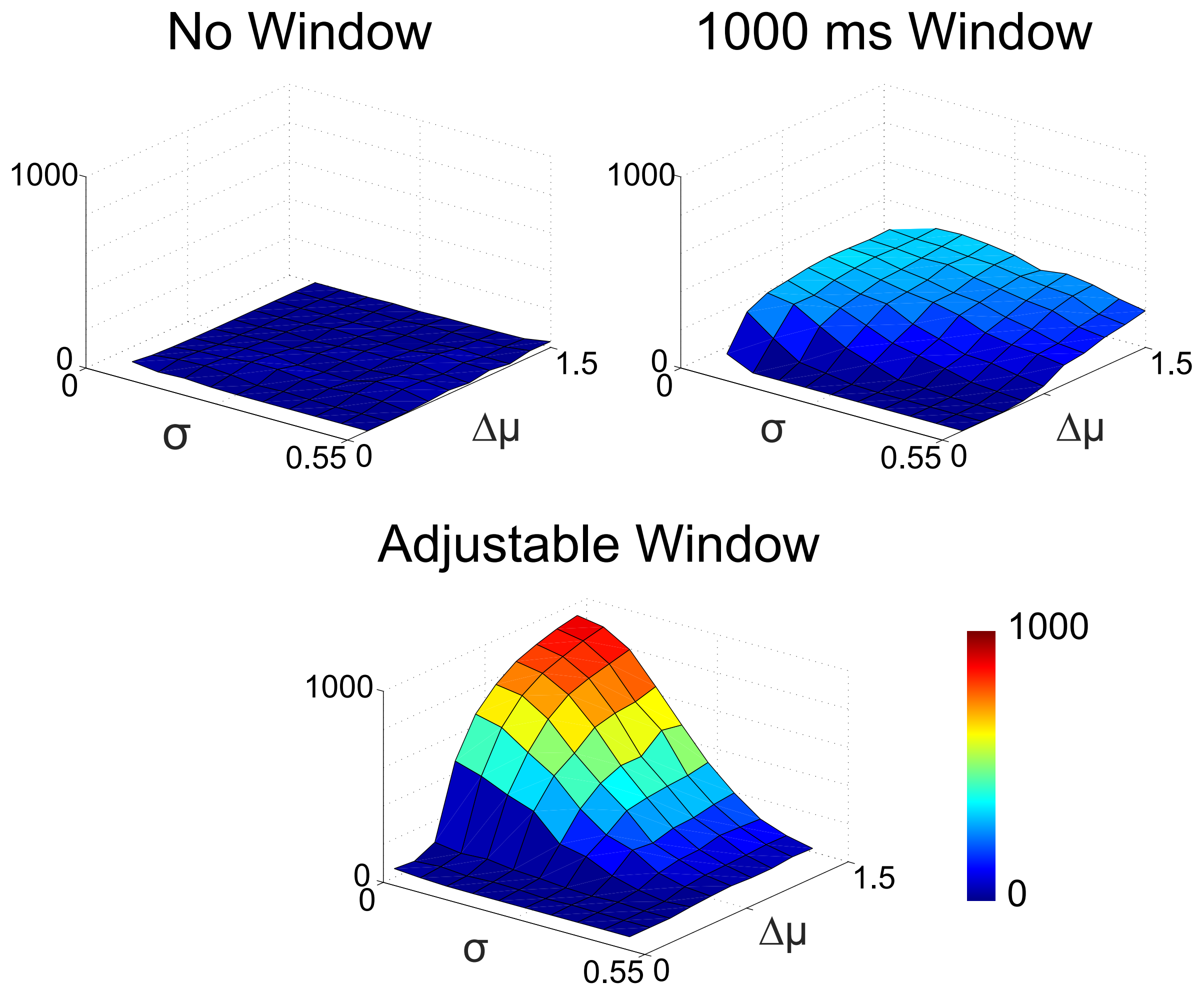


Figure 3.7: Average cumulative reward the fourth epoch (3000-4000ms) of fig. 3.4; average of 1000 runs. Top left, top right and bottom correspond to algorithm with no window, fixed (1000ms) window, and adjustable window, respectively

Analysing the input stimuli used so far, one can argue that they are not very ethologically plausible or realistic. We wouldn't expect that in a real task the new options presented to the decision maker would always be of increased mean when compared to its preceding winning option, or that all options would keep their mean constant until the end of the trial.

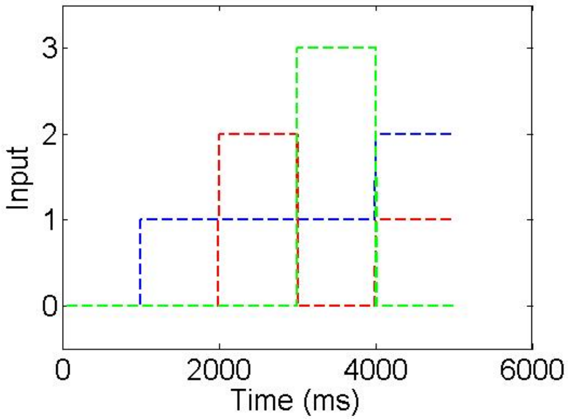


Figure 3.8: Unconstrained mean inputs with 4 competing options; means for each option are displayed in different colours.

In order to show the robustness of the algorithm in more realistic scenarios, we designed the set of inputs in 3.8; these inputs have a number of characteristics which are present in real world decisions. The first of these characteristics is the ability to change means more than once over the whole trial, another is having inputs that appear later in the trial, disappear and then reappear as competition for the winning option. Finally we introduced a type of option that appears as the winning option and disappears after one epoch.

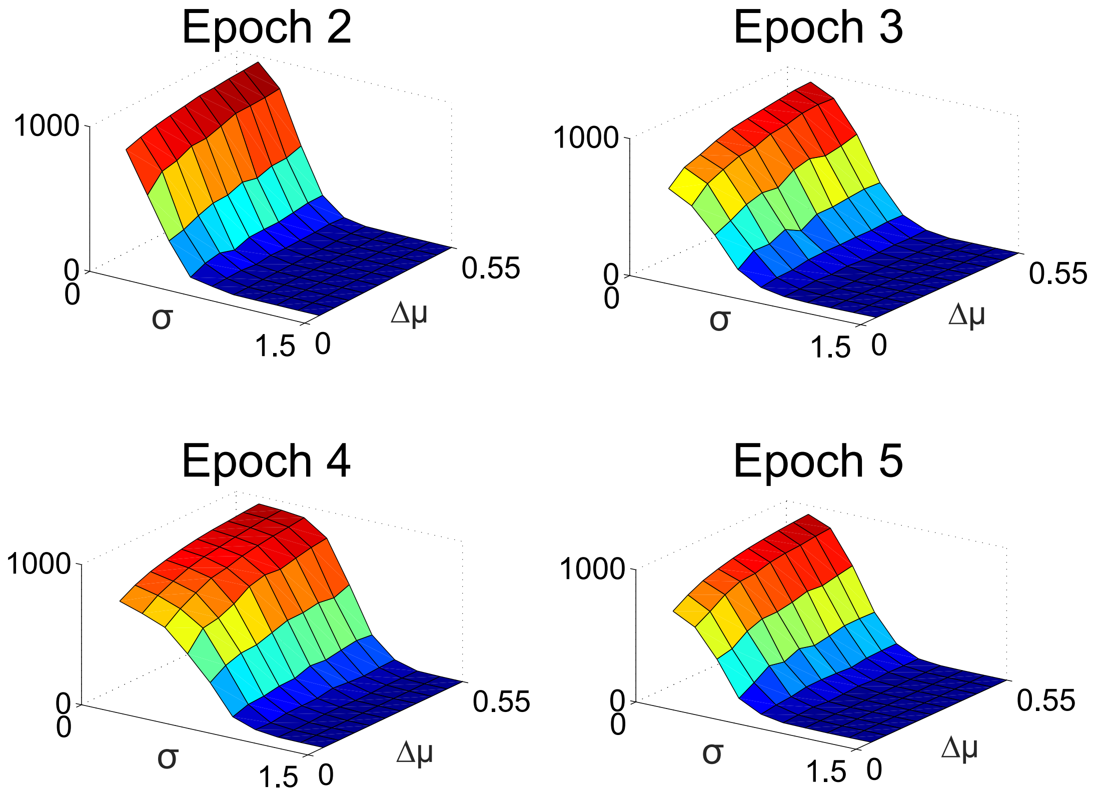


Figure 3.9: Average cumulative reward for each epoch of fig. 3.8; average of 1000 runs. Row 1 column 1 corresponds to second epoch (1000-2000 ms), row 1 column 2 corresponds to third epoch (2000-3000 ms), row 2 column 1 corresponds to fourth epoch (3000-4000 ms), row 2 column 2 corresponds to fifth epoch (4000-5000 ms)

As can be observed in fig. 3.9, the reward levels are very similar to those in figs. 3.5 to 3.7 also display a steady shape of the reward curve throughout the trial. These results along with the previous ones indicate that the addition of a rectangular window dependent on the difference between signal means is an efficient way of extending a traditional MSPRT to accommodate different types of non-stationary evidence.

Chapter 4

Application of algorithm to shepherding-flock environment

The behaviour of flock of sheep has long held the interest of researchers. For decades, a large body of work has stemmed from attempting to understand flock dynamics and behaviours particular to this class of animals (Hunter and Milner, 1963; Arnold and Pahl, 1974; Arnold et al., 1981; Vaughan et al., 2000; King et al., 2012).

Despite the longstanding and extensive literature describing different flock behaviours, only more recently have those behaviours been modeled through computer simulations (Vaughan et al., 2000; King et al., 2012; Strömbom et al., 2014).

One of the more comprehensive models of sheep behaviour was presented by Strömbom et al. (2014); in this publication, the authors describe how they observed the interactions between sheepdog and merino sheep in the south of Australia, recording one herding event per day. In each event, the shepherd dog was directed to herd the flock with minimal guidance.

The computational model developed by the authors aimed to mimic the behaviours observed and was based on a self-propelled particle model of local attraction-repulsion.

In the current chapter, we will take an existing model of shepherd-sheep behaviour based on observation of those two groups of animals and embed our full MSPRT-like algorithm as the shepherd's decision mechanism.

The main advantage of this implementation is that, since the sheep behaviour was modeled based on real flock interactions observed, this constitutes a close approximation of a real shepherding scenario and therefore presents an ethologically plausible context in which to test our algorithm's capabilities.

4.1 Shepherd and Sheep agents

The actions performed by the shepherd at each time step will depend on the flock's cohesion, given by the positions of the sheep relative to the General Centre of Mass (GCM). These actions and the positions adopted by the shepherd (P_d and P_c) when performing each action are illustrated in fig. 4.1, from Strömbom et al. (2014).

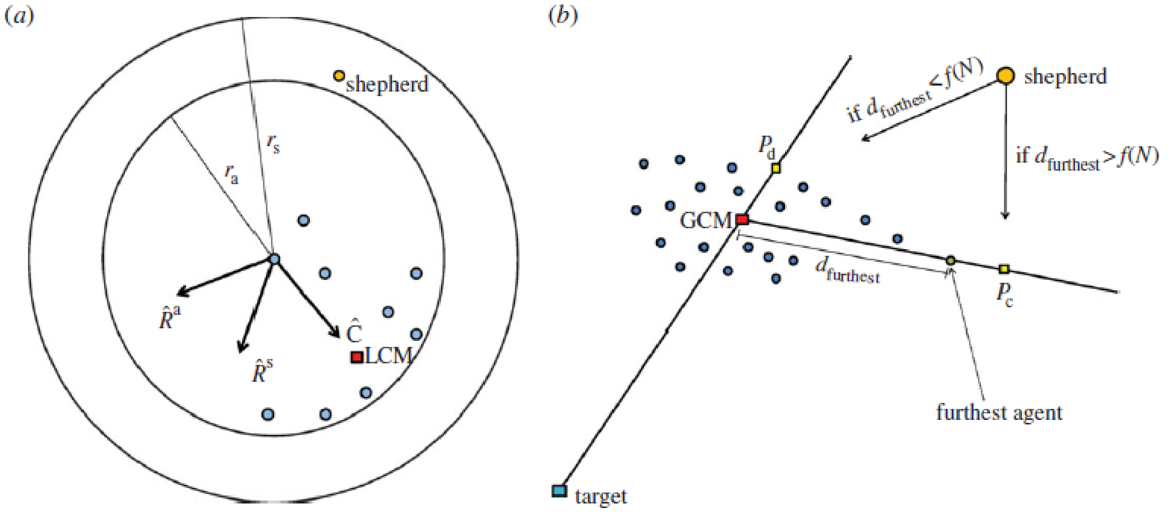


Figure 4.1: Interaction rules between shepherd and sheep (from Strömbom et al. (2014)).

In Strömbom et al. (2014), when the algorithm is initialised, N sheep are placed in random positions in the upper right quarter of an L^*L field and one shepherd is placed in a random position in any of the remaining quarters. The shepherd's task is to guide the flock to the lower left corner of the square, making sure that the flock's centre of mass (GCM) is within $f(N)$ of the origin, $f(N)$ will vary according to the number of sheep in the flock and is given by $f(N) = r_a N^{2/3}$, where r_a is the distance, between the shepherd and the closest member of the flock, at which the shepherd reduces its speed .

The shepherd herds the agents by performing one of two actions: collect out lier or herd the flock, this can also be done at different speeds depending on the proximity to any member of the flock.

The shepherd will go into 'collect' mode if one or more agents are not within $f(N)$ of the GCM; for this, it will adopt a collecting position P_c, behind the agent which is furthest from the GCM. When all agents are cohesive, at a distance smaller than $f(N)$ from the GCM, the shepherd will switch to 'driving' mode, adopting the driving position P_d. This position will be directly behind the flock, with relation to the target.

Once in one of the two modes (collect or drive), the shepherd's speed will be adjusted according to the its distance to the closest agent. If all agents are at a distance greater than $3r_a$ from the shepherd, then it will move at a speed of δ_s. If any agent is at a distance smaller or equal to r_a, then the shepherd will adjust its speed to $0.3r_a$.

It should be noted that no planning is involved in this algorithm and that the shepherd simply computes the evidence available at each point in time and acts if this has crossed the threshold; shepherd behaviour is illustrated in fig. 4.2.

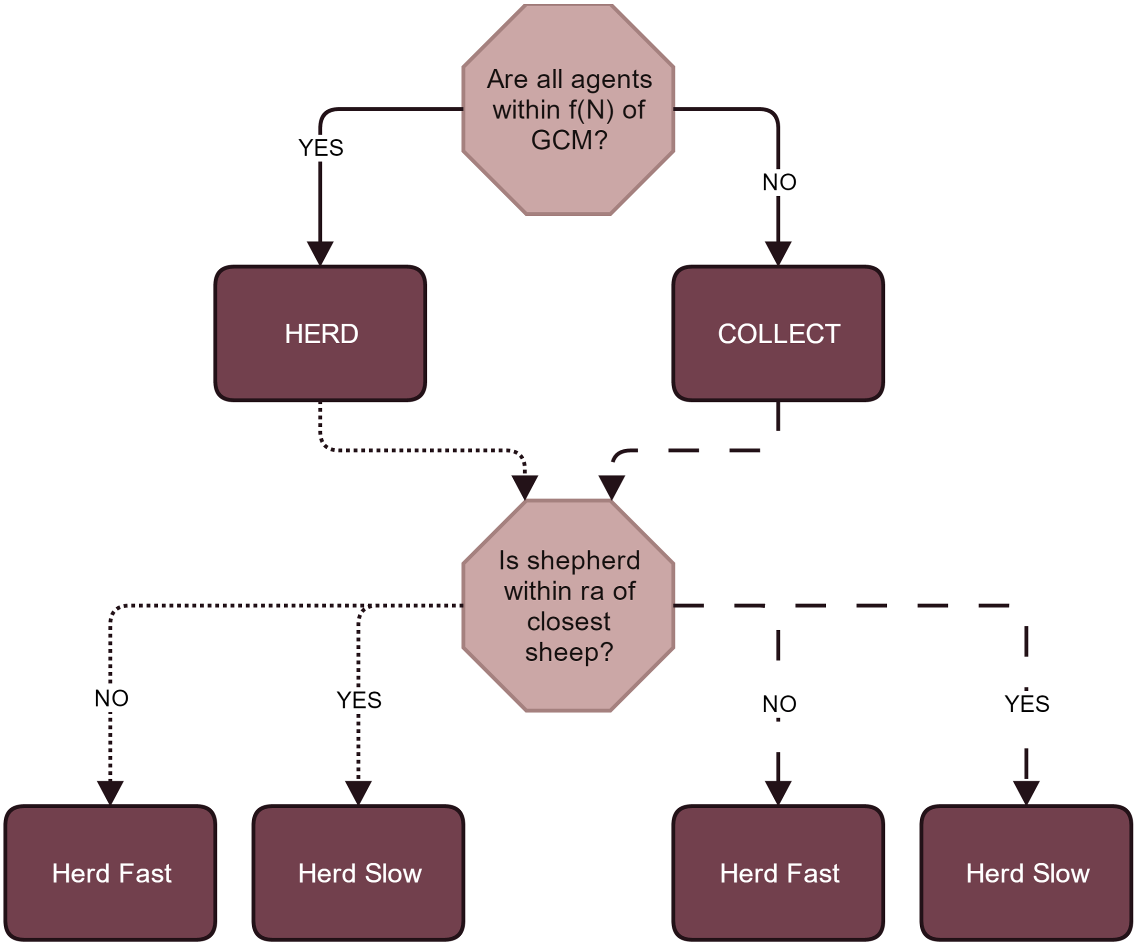

Are all agents within f(N) of GCM?
YES
NO
HERD
COLLECT
Is shepherd within ra of closest sheep?
NO
YES
NO
YES
Herd Fast
Herd Slow
Herd Fast
Herd Slow

Figure 4.2: Flowchart of the algorithm describing Shepherd Behaviour; each action performed by the shepherd will depend on the sheep's individual and collective centres of mass.

A well known principle of group behaviour is known as'selfish-herd theory' (Krause and Ruxton, 2002), this theory dictates that more cohesive groups are better at detecting predators than more disperse ones. As such, individuals in a flock should demonstrate a tendency for attraction to their closest neighbours in order to reduce the risk presented by the predator. This behaviour is present in this model, where the sheep demonstrate a tendency to remain in the proximity of their n closest neighbours, unless their distance is smaller than r_a, in which case they will repel each other. For agent i, at position $\overline{A}_i$ with k neighbours at positions $\overline{A}_1, ..., \overline{A}_k$, the strength of inter-agent repulsion R_i^a is given by

$$\overline{R}_i^a = \sum_{j=1}^{K} \frac{1}{|\overline{A}_i - \overline{A}_j|}(\overline{A}_i - \overline{A}_j)$$

.

Sheep are also attracted to the centre of mass of their n nearest neighbours $L\overline{C}M_i$ in the direction of

$$\overline{C}_i = L\overline{C}M_i - \overline{A}_i$$

When the distance between a sheep and the shepherd, at position $\overline{S}$ is smaller than r_s, the sheep is repelled directly away from it in the direction of

$$\overline{R}_i^s = \overline{A}_i - \overline{S}$$

If this distance is larger than r_s the agent will not be repelled by the shepherd and will instead exhibit small random movements. These are important to resolve situations of deadlock since they add a small amount of noise to the simulation.

The sheep's heading at a particular point in time is $\overline{H}_i^{`}$ a weighted combination of all the forces described previously, with the addition of inertia and error terms ($\hat{H}_i^{\prime}$ and $\hat{\varepsilon}$) and parameters h, c, ρ_a, ρ_s, and e described in Table 4.1; this is given by :

$$\overline{H}_i^{`} = h\hat{H}_i + c\hat{C}_i + \rho_a\hat{R}_i^a + \rho_s\hat{R}_i^s + e\hat{\varepsilon}_i$$

Sheep behaviour is illustrated in fig. 4.3:

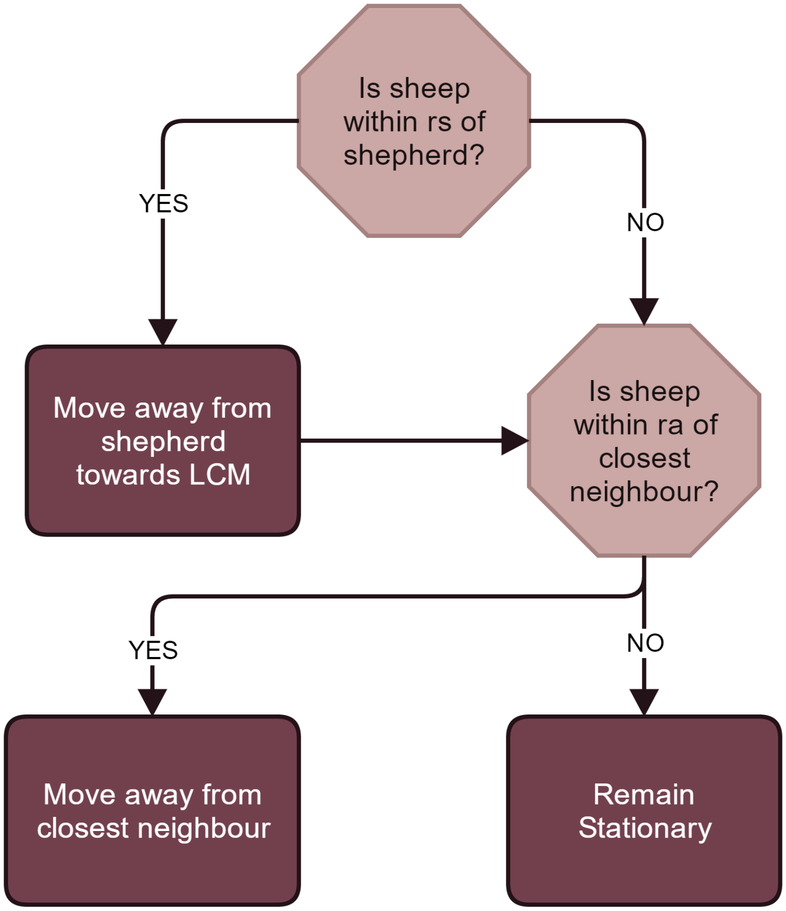

Is sheep within rs of shepherd?
YES
NO
Move away from shepherd towards LCM
Is sheep within ra of closest neighbour?
YES
NO
Move away from closest neighbour
Remain Stationary

Figure 4.3: Sheep Behaviour.

4.1.1 Ethologically plausible evidence

In the previous section we looked at the behaviour of both types of agents, shepherd and sheep; these were described by a series of group and individual behaviours of repulsion and attraction.

As described, the behaviour of the shepherding agent was straightforward since it had perfect knowledge of sheep positions and distances to itself and the GCM. As such, from a decision making standpoint, the decision maker (shepherd) used only one sample of noise-free evidence to decide which action to perform next. As soon as one of the sheep strayed from the flock, the shepherd could collect it immediately, quickly resuming flock cohesion. Its behaviour was described in the following way:

- Shepherd has access to position of each sheep
- Center of mass of flock calculated
- Shepherd identifies closest sheep from it and sheep that is furthest from center of mass.
- Shepherd decides whether to herd flock or collect outlier.

This approach to shepherd behaviour becomes increasingly unrealistic with the growing number of sheep, since despite the increased difficulty of the task itself (more agents to control), the difficulty of decision is maintained since the shepherd knows exactly where each of the sheep are.

A more realistic shepherd behaviour would be the following:

- Shepherd observes the field
- Noisy position of each sheep is estimated
- Center of mass of flock is estimated based on noisy position of agents
- Shepherd estimates closest sheep from it and sheep that is furthest from center of mass.
- Probability of performing each action (Herd quick and slow, collect quick and slow) is calculated based on sheep distance to shepherd and GCM
- Previously calculated probabilities are used as inputs
- Accumulated inputs are used to reach a decision.

4.2 Methods

When deciding what action to take the shepherd should act in two steps, firstly, decide whether to collect or herd, and secondly at what speed these actions should be taken. For

$$D_{max}^{GCM} = MAX(D_i^{GCM}), D_{min}^{S} = MAX(D_i^{S})$$

Where D_{max}^{GCM} is the maximum distance from a sheep to the centre of mass of the flock, and D_{min}^{S} is the minimum distance from a sheep to the shepherd.

If $D_{min}^{S} \leq 3 * r_a$, $\delta_s = 0.3 * r_a$ - the shepherd should slow down to maintain distance from sheep.

If $D_{max}^{GCM} < f(N)$, the flock is cohesive and the shepherd should position itself behind it and move in the direction of the target.

If $D_{max}^{GCM} > f(N)$, there is at least one outlier and the shepherd should herd this sheep towards the GCM.

Parameter values for the algorithm can be found in table 4.1.

Figure 4.5 shows a typical trial, where the black dot represents the shepherding agent and the white dots represent the sheep. It illustrated both the behaviours of repulsion from shepherd (2,3 and 4) and attraction to neighbours (3 and 4) mentioned previously.

PARAMETER	VALUE	DESCRIPTION
L	150	Length of square field
N	2-150	Number of agents - sheep
n	1-149	Number of nearest neighbours -sheep
r_s	75	Shepherd detection distance
r_a	1	Agent to agent interaction distance
ρ_a	2	Relative strength of repulsion from other sheep agents
c	1.2	Relative strength of attraction to n nearest neighbours
ρ_s	1	Relative strength of repulsion from shepherd
h	0.5	Relative strength of proceeding in the previous direction
e	0.1	Relative strength of angular noise
δ	1	Agent displacement per time step
p	0.05	probability of moving per time step while grazing
σ_p	1	Standard deviation of Gaussian noise of sheep position
δ_s	$1.5 * \delta$	Shepherd displacement per time step
P_d	$r_a * \sqrt[2]{N}$	Driving position behind the flock
P_c	r_a	Collecting position behind furthest agent
m_{time}	8000	Maximum number of time steps for a successful simulation

Table 4.1: Model parameters

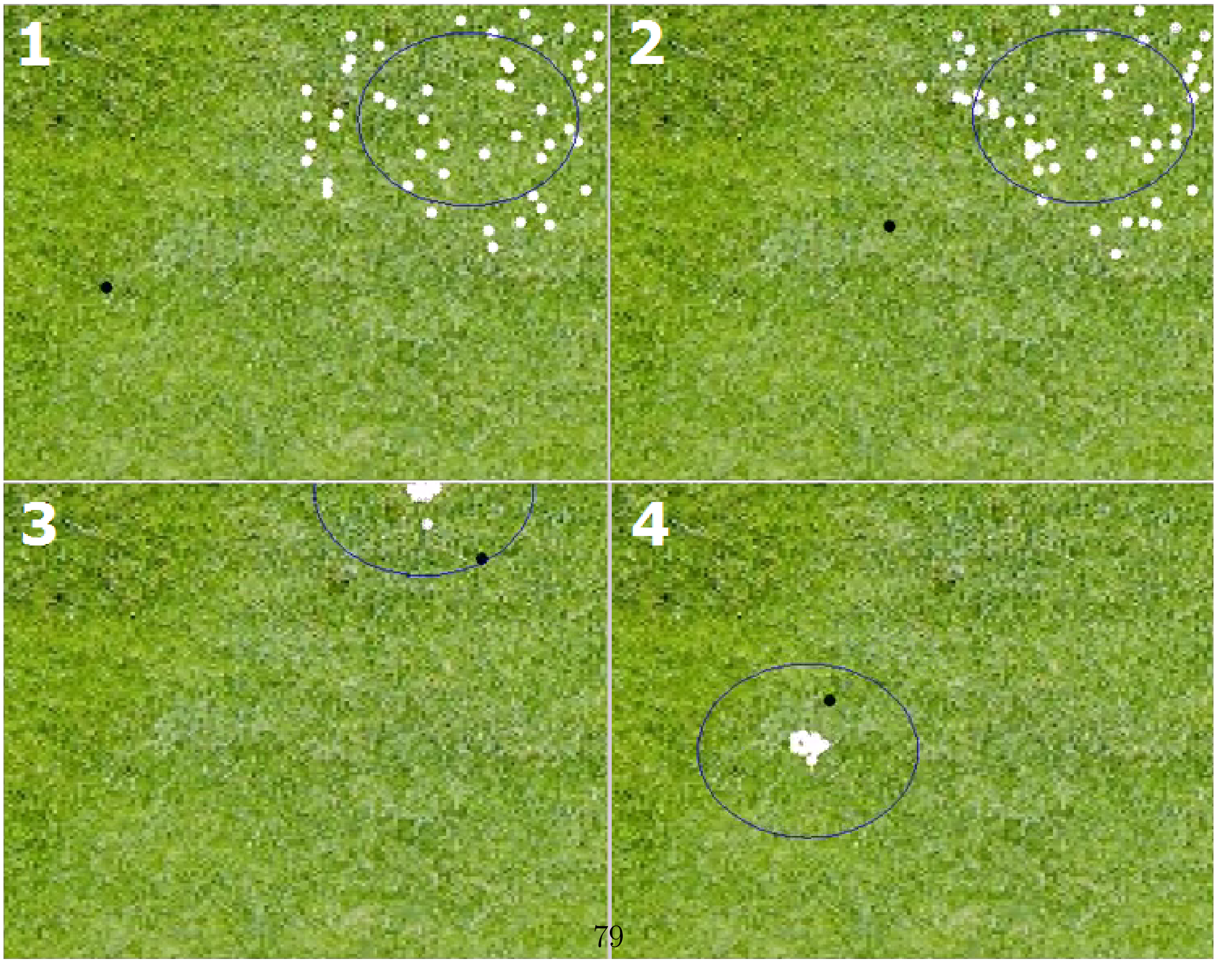


Figure 4.4: Typical shepherding trial illustrating repulsion from shepherd (2,3,4) and attraction to neighbours (3 and 4).

In Strömbom et al. (2014), decisions were computed in one time step since the shepherd had perfect information about the location of the sheep.

Since we were interested in exploring whether there was an advantage to the use of the MSPRT in place of the existing shepherd model, we redesigned the experiment in a way which allowed us to get a grasp on the effectiveness of our model. For this, 4 different actions were defined:

- Herd Fast (C_f) - $D_{max}^{GCM} < f(N)$ and $D_{min}^{S} > 3r_a$
- Herd Slow (C_s) - $D_{max}^{GCM} < f(N)$ and $D_{min}^{S} < 3r_a$
- Collect Fast (H_f) - $D_{max}^{GCM} > f(N)$ and $D_{min}^{S} > 3r_a$
- Collect Slow (H_s) - $D_{max}^{GCM} > f(N)$ and $D_{min}^{S} < 3r_a$

These actions were modeled as probabilities resulting from a combination of the estimated values of two measures of distance; smallest distance to shepherd and largest distance to GCM.

Through the estimation of the distance (d) of the closest agent to shepherd, and of furthest agent to GCM (in this case with 100% accuracy due to the lack of noise), we can compute the probability of these being over the distance thresholds, $3 + r_a$ and $f(N)$, this is done using the Z score with (S is the standard deviation):

$$z = \frac{d - d_{th}}{S}$$

The probability for each distance was then given by the complementary error function:

$$p = \frac{2}{\sqrt(\pi)} \int_{d}^{\infty} e^{-t^2} dt$$

Given the probabilities obtained for D_{min}^{S} and D_{max}^{GCM}, the action's probability were obtained by:

$$P_{H_f} = (1 - P_{D^{GCM}_{max}}) * P_{D^{S}_{min}}$$

$$P_{H_s} = (1 - P_{D^{GCM}_{max}}) * (1 - P_{D^{S}_{min}})$$

$$P_{C_f} = P_{D^{GCM}_{max}} * P_{D^{S}_{min}}$$

$$P_{C_s} = P_{D^{GCM}_{max}} * (1 - P_{D^{S}_{min}})$$

These were fed into the MSPRT algorithm with a fixed threshold of 0.5, and a decay parameter $\lambda = 0.3$; these parameters were empirically derived. In order to allow for a meaningful comparison to the original shepherd, the noise parameter was set to 0; the MSPRT decision maker was then considered to have perfect knowledge.

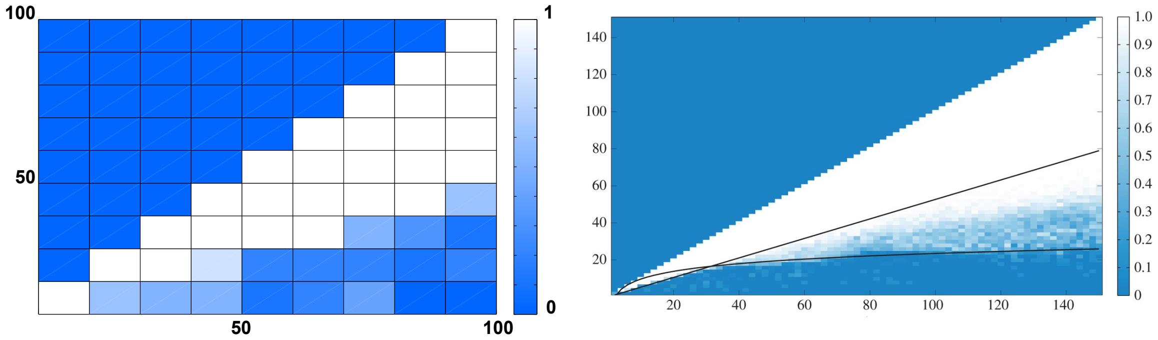


Figure 4.5: Proportion of successful shepherding events for a number of agents (N - x axis) vs a number of neighbours (n - y axis). Left panel shows result of our algorithm, right panel taken from Strömbom et al. (2014)

A shepherding event is considered successful when the shepherd places the center of mass of the flock in the target in under 8000 time steps. Results from both noise-free implementations can be observed in fig. 4.5; there, successful shepherding is shown to occur roughly between $n = N - 1$ and and $n = 0,53N$, with the probability of success dropping after this point, for both cases.

Strömbom et al. (2014) argue that the algorithm developed is highly paralleled with real sheep-dog interactions, making it biologically plausible; therefore, by approximating the results from it, we argue that our implementation can also be considered biologically plausible.

After establishing the parallel between both implementations, the next logical step was to break the input evidence constraints through the addition of noise. As a first step, we added noise to the sheep's positions so that if the sheep is at (x, y), the shepherd would perceive the sheep to be within $x \pm 1, y \pm 1$. Therefore, the euclidean distance to shepherd or GCM could vary ± 1.42 from the real value. Also, since the GCM was calculated using the noisy positions of the sheep, this variable would be intrinsically noisy.

4.2.1 Combination of MSPRT & SPRT

At each time step, the shepherd is presented with two possible decisions:

- **Are the sheep cohesive?**

  If yes: HERD

  If no: **Which sheep should be collected first?** COLLECT

Due to the nature of the decisions to be performed by the shepherd, one would expect a single or double decision process, depending on the result of the first step.

Initially, the shepherd has to decide whether all sheep are within a certain distance from the GCM. Since each sheep's distance is independent from the others, determining whether or not the sheep are cohesive will require N decisions, with N being the number of sheep. If the sheep are cohesive, the shepherd must herd them to the target, if not, it must decide which sheep to collect first based on their individual distances to the GCM.

For our implementation of this mechanism we designed a two step decision process based on a combination of the SPRT ($N_h = 2$ competing hypotheses) and the MSPRT ($N_h > 2$ competing hypotheses). For this, N SPRT computations (one for each sheep) were performed, with a subsequent MSPRT computation based on the result of the first step.

Step 1 - SPRT:

- $H_{zn} = (d_i > r_o)$; with d_i = distance from sheep to center of mass, r_o = zone limit, for n^{th} agent

H_{zn} states that the nth agent is outside of the zone limit, therefore constituting an out lier to the group.
If all H_{zn} are false, all sheep are inside the center of mass zone and the shepherd must adopt the driving position.
If any H_{zn} is true, the shepherd must find the out lier by performing step 2.

The use of the SPRT algorithm entails similar challenges to the MSPRT since, like its multi-alternative counterpart, it is not equipped with a mechanism to discard outdated samples of evidence. Due to the focus of our work on the MSPRT, the SPRT algorithm was implemented with a fixed integration window of size 100 and an interrogation paradigm (forced decision) in order to simplify parameterisation. The interrogation paradigm worked by calculating the number of time steps since the last decision occurred and forcing a new one once a set number was reached. Three different forced decision times were used; 1,10, and 100.

Step 2 -MSPRT:

- $H_i = (d_i = max\{d_i\})$

H_i states that sheep i is the furthest from the centre of mass of the flock. d_i is the noisy measure of distance of agent i to centre of mass.

Unlike in our initial chapter where we used the percentage of successful shepherding events as our only measure of result quality, because we are interested in optimising both speed and accuracy, for the following results we also examined average time to target, which was defined as the average time it took the shepherd to successfully herd the agents to the target. For parameter sets where no events were successfully completed, the time to decision was considered 8000 steps (maximum amount of time steps in a successful event).

The advantage of including time to target in our analysis is the potential it presents to disentangle speed and accuracy, which cannot be done when observing the % of successful events, since this reflects a combination of both. A parameter set with a higher % of successful events should not necessarily be favoured over one with a slightly lower percentage if the latter has a significantly lower decision time; depending on the type of application of the algorithm, one might chose to abdicate of some accuracy in favour of speed.

Despite the use of a forced decision mechanism in the first step of this decision process, the second step (MSPRT) does not have a mechanism for forcing a decision; for this reason, it is possible for deadlocks to occur, especially in situations where the evidence is changing rapidly or two agents are at a very similar or even the same distance from the centre of mass, since decisions are only performed when evidence for one of the options crosses the threshold.

Table 4.2 shows the result of the preliminary simulations for the SPRT+ MSPRT

Time to Decision of SPRT	Average % of Successful Shepherding Events	Average time to target without failed trials	Average time to target with failed trials
1	87	2157	2823
10	90	2131	2634
100	87	2029	2731

Table 4.2: Table contains the average % of successful shepherding events and average time to target for number of different 'time to decision of SPRT'. MSPRT was implemented with a 0.01 exponential window. Results from full algorithm with N=10-100 sheep and 9-99 neighbours, average of 55 sets of parameters (data points), with 50 simulations for each data point. Fourth column shows average time to target of completed events (i.e. if only one event was completed out of 100 with a time of 2000, that is the value displayed); fifth column shows average time to target for all events, time accounted for non-completed targets is 8000

paradigm. From the second column, which displays the percentage of successful events, one can note that the 10 step forced decisions performed the best out of the three parameters, with a 3% difference from the other implementations. Column 3, which displays the average time to target as defined previously, on the other hand, shows that a 100 step forced decision performs best, with a 5% and 6% difference to 10 and 1 steps, respectively.

This set of results offers an interesting insight into the different dynamics that are present in this algorithm, with the results contradicting the initial expectations. If we take into account solely the 'time to decision of the SPRT', one might expect this value to be directly proportional to the average time to target, since this allows the shepherd to act faster by increasing the frequency at which decisions are made. Unlike what was expected, with the relationship between the aforementioned parameter and the 'average time to target without failed trials' was in fact one of inverse proportionality.

Setting the SPRT step of the algorithm to function with 1 time step turns the algorithm into a single-sample decision process, which is more prone to error since it does not average out the step by step variability of the inputs. This setting is the one used by Strömbom et al. (2014) in their implementation of the shepherd, which assumed perfect knowledge of the sheep's positions. Unlike the Strombom model, our implementation is subject to noise and as such is increasingly prone to errors with the decrease in number of samples.

Despite the relationship demonstrated between columns 1 and 3, no such inference can be made regarding column 2, where the average % of successfull shepherding events bares no linear relationship to the time to decision of the SPRT.

The MSPRT algorithm implemented in this decision mechanism was that described in previous chapters where a window of varying size was used to discard outdated evidence. The rectangular window size and corresponding λ parameter for the exponential window are described in equations 4.1 and 4.2, respectively.

In order to display the time to decision in context with the number of successful shepherding events, the time to target with failed trials was also included; this is calculated taking into account simulations which weren't finished as having a time to target of 8000.

$$DT(\Delta\mu) = 445.7\Delta\mu^{-1.98} \tag{4.1}$$

$$DT(\lambda) = 1/DT(\Delta\mu) \tag{4.2}$$

In order to understand whether there was a relevant relationship between 'time to decision' and the measures of accuracy and speed, we decided to recreate the conditions tested throughout our work where a number of exponential and rectangular windows were implemented. For this, we used 5 different windows of two types: two rectangular windows of size 10 and 100 and three exponential windows with λ of 0.01,0.05, and 0.1.

Results for the rectangular windows can be observed in table 4.3; each window length determines the number of latest samples of evidence that will be used to compute a decision. For this window geometry, there was a slight difference in '% of successful shepherding events' between different 'time to decision' for the 10 step window, (3% at the most), but no difference for the 100 step window.

window size	Time to decision of SPRT	% successful shepherding events	Average time to target without failed trials	Average time to target with failed trials
10	1	49	2302	5218
10	10	47	2220	5274
10	100	46	2224	5392
100	1	86	2115	2828
100	10	86	2019	2818
100	100	86	2095	2837

Table 4.3: Table contains the average % of successful shepherding events and average time to target for number of different 'time to decision of SPRT' with rectangular windows of different sizes. Results from full algorithm with N=10-100 sheep and 9-99 neighbours, average of 55 sets of parameters (data points), with 50 simulations for each data point. Fourth column shows average time to target of completed events; fifth column shows average time to target for all events, time accounted for non-completed targets is 8000.

Surprisingly, unlike the preliminary simulations where the 'time to decision of SPRT' was inversely proportional to the 'time to target without failed trials', with the rectangular window paradigm these two quantities were proportional, indicating that for a window of rectangular geometry fewer decisions lead to increased trial length. It is worth noting that no relationship was found between 'time to decision of SPRT' and 'time to target without failed trials'.

Table 4.4 shows the results for windows of exponential geometry, where the λ parameter determines the rate of decay of evidence. The first 3 rows in the table are repeated from table 4.2.

Starting on the second row ('% of successful shepherding events') comparing this value for each λ with the same 'time to decision of SPRT', the trend is for this percentage to decrease with the increase of λ, this is to be expected since an increase in this parameter leads to a quicker discard of evidence and subsequently to a smaller amount of evidence being used to compute decisions. The decrease in amount of evidence can lead to inaccurate decisions since the effect of the noise in each sample is amplified.

λ	Time to decision of SPRT	% successful shepherd-ing events	Average time to target without failed trials	Average time to target with failed trials
0.01	1	87	2157	2823
0.01	10	90	2131	2634
0.01	100	87	2029	2731
0.05	1	61	2366	4585
0.05	10	57	2248	4725
0.05	100	58	2016	4529
0.1	1	52	2449	5090
0.1	10	54	2460	4980
0.1	100	55	2188	4844

Table 4.4: Table contains the average % of successful shepherding events and average time to target for number of different 'time to decision of SPRT' with exponential windows of different sizes. Results from full algorithm with N=10-100 sheep and 9-99 neighbours, average of 55 sets of parameters (data points), with 50 simulations for each data point. Fourth column shows average time to target of completed events ; fifth column shows average time to target for all events, time accounted for non-completed targets is 8000.

In line with the results in table 4.3, no meaningful relationship was found between the 'time to decision of SPRT' and the '% of successful shepherding events for $\lambda = 0.01$.

Further, with regards to the relationship between 'Average time to target without failed trials' and 'time to decision of SPRT' these were found to be inversely proportional, with the former decreasing with the increase of the latter. These results were consistent across the range of λ and 'time to decision of SPRT'. No proportionality relationship was found between 'Average time to target with failed trials' and 'time to decision of SPRT'.

Due to the reduction in decision time from the increase of the parameters in column two (both with and without accounting for failed trials), and the fact that there is no consistent relationship between this and the results in column 3, we decided to set the the 'time to decision of SPRT' as 100 for all subsequent simulations; this allowed us to decrease the length of the simulation thereby increasing the accuracy of decision with negligible cost to the percentage of successful events.

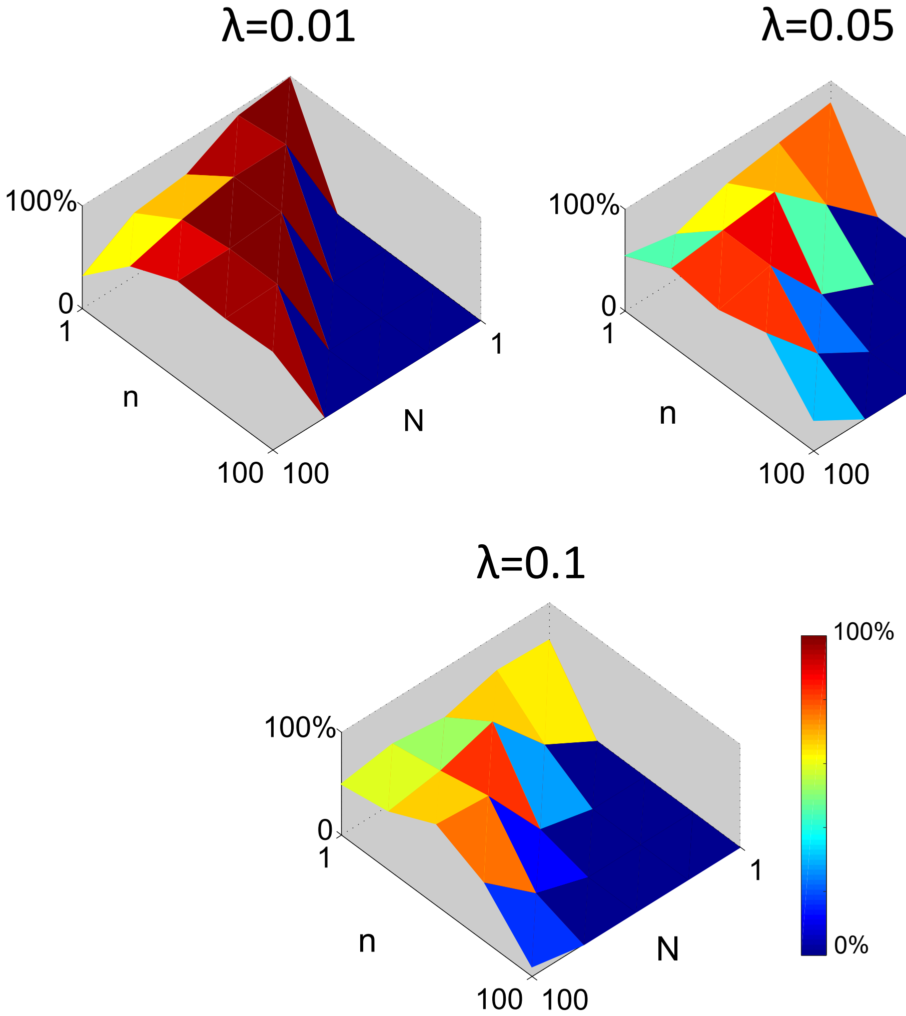


Figure 4.6: Percentage of successful shepherding events for a number of agents (N) vs a number of neighbours (n). Each panel shows this percentage for each set of parameters with the exponential decay factor λ - 50 trials for each data point.

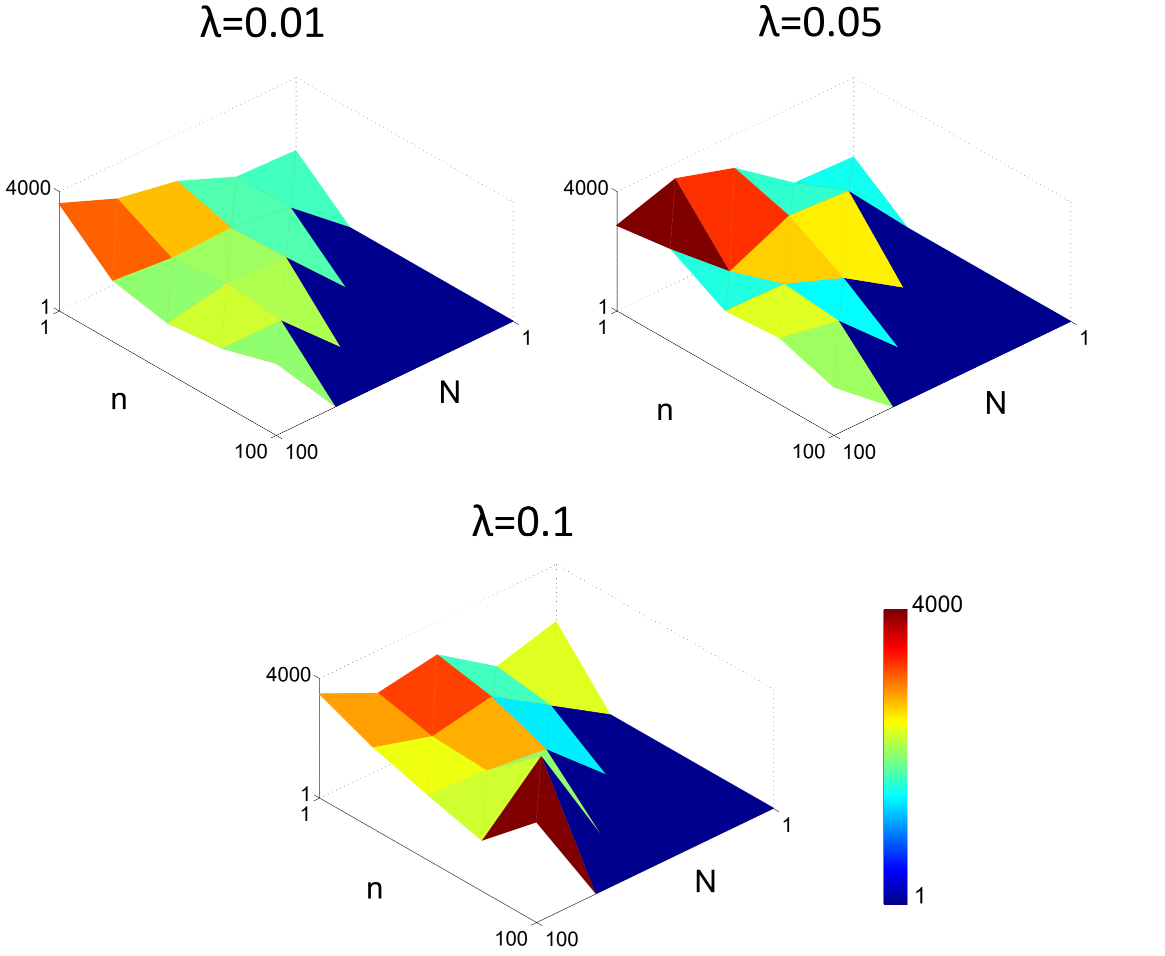
Figure 4.6 displays the percentage of successful shepherding events for a number of agents (N) vs a number of neighbours (n) for each value of λ, with the 'time to decision' fixed at 100. These figures are important since they allow for a much deeper understanding of the dynamics of the simulation since they are a less condensed version of results.

In the aforementioned figure, the lighter colours indicate higher percentages of successful shepherding events and the darker colours lower ones. The first trend to be noted between figures is that the increase in λ leads to a decrease in correct selection, in line with the high level results in table 4.4.

Also common to all figures is that the best results lay around n=N-1, which depicts the case where the number of neighbours is close to the number of agents, i.e. the herd has a strong tendency to maintain group cohesion, making it easier for the shepherd to guide the agents to the target without these scattering through the field.

For a fixed number of agents N, as we decrease n, the percentage of successful shepherding events also decreases; this is due to the fact that the agents will be attracted to fewer of their neighbours, leading to the separation of the main group into smaller groups and making it more difficult for the shepherd to herd the agents to the target without having to constantly collect outliers.

Based on the previous results, one would expect the 'average time to target' to increase as the difference between N and n increases.



λ=0.01

λ=0.05

λ=0.1

Figure 4.7: Average time to target without failed trials for a number of agents (N) vs a number of neighbours (n). Each panel shows this time for each set of parameters with the exponential decay factor λ - 50 trials for each data point.

Figure 4.7 displays the average time to target for a number of agents (N) vs a number of neighbours (n) for each value of λ, with the 'time to decision' fixed at 100.

As predicted, the time to target increases as $n << N-1$. These results also follow those in fig. 4.6 where $\lambda = 0.01$ resulted in the lowest decision times overall. The peaks in the graphs, especially noticeable for $\lambda = 0.05$ and $\lambda = 0.1$ are due to the simulations timing out for those parameter sets, which pushed the average up (since these are considered as maximum time).

Despite the variability in time for each λ, the overall performance is considered positive since all data points are under 4000, which is half of the maximum time before time out.

Previous results are in line with previous chapters where the optimal balance between speed and accuracy was shown to lay at or around $\lambda = 0.01$.

4.2.2 MSPRT vs biologically accurate shepherd

Strömbom et al. (2014) argued for the ethological validity of their model given that it was based on observation of a real shepherd dog's behaviour. However good it was at predicting and acting on sheep behaviour, the shepherd presented by the authors was highly constrained to situations of absolute knowledge of the positions of each member of the herd and therefore not very realistic.

Part of the reasoning behind the application of the MSPRT as the decision making mechanism for the shepherd was the hypothesis that the existing model would only perform as expected using the parameters defined in Strömbom et al. (2014). We believed that when in a similar environment to that used to model the simulations in fig. 4.6, the shepherd would struggle to cope with the uncertainty in the positions of the agents and therefore under perform when compared to the more robust MSPRT.

In order to test this idea, we replicated this 'biologically accurate' shepherd agent, with an additional mechanism which allowed us to test it in noisy environments. For this, we used the same model, but changed the inputs to the shepherd so that its perception of sheep position was affected by noise. This also led to a noisy estimation of their general centre of mass.

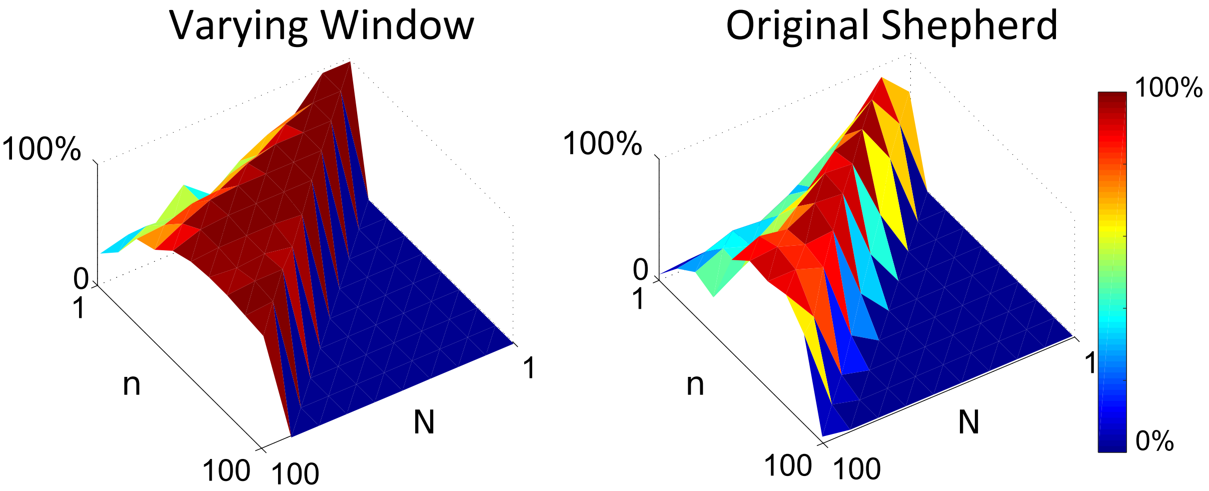


Figure 4.8: Percentage of successful shepherding events for a number of agents (N) vs a number of neighbours (n). Left panels shows this percentage for the SPRT+MSPRT shepherd and right panel for the Strömbom et al. (2014) shepherd

Figure 4.8 shows the percentage of successful shepherding events for a number of agents (N) vs a number of neighbours (n). Left panels shows this percentage for the SPRT+MSPRT shepherd and right panel for the Strömbom et al. (2014) shepherd.

When comparing the results in both panels, it can be noted that these display similar patterns of results, with the highest percentages of success occurring for $n = N - 1$, however, the original shepherding algorithm shows an accentuated drop for $n < 0.7 * N$, while the MSPRT shepherd maintains percentages over the majority of the range, only dropping below 50% accuracy for 4 out of 55 sets of parameters, where $n << N$

On average, over the entire parameter range, the MSPRT shepherd had an 84% success rate, whereas the original shepherd was limited to a 55% rate. This represents a 1.5x increase in successful shepherding events resulting from the addition of the MSPRT as a decision mechanism.

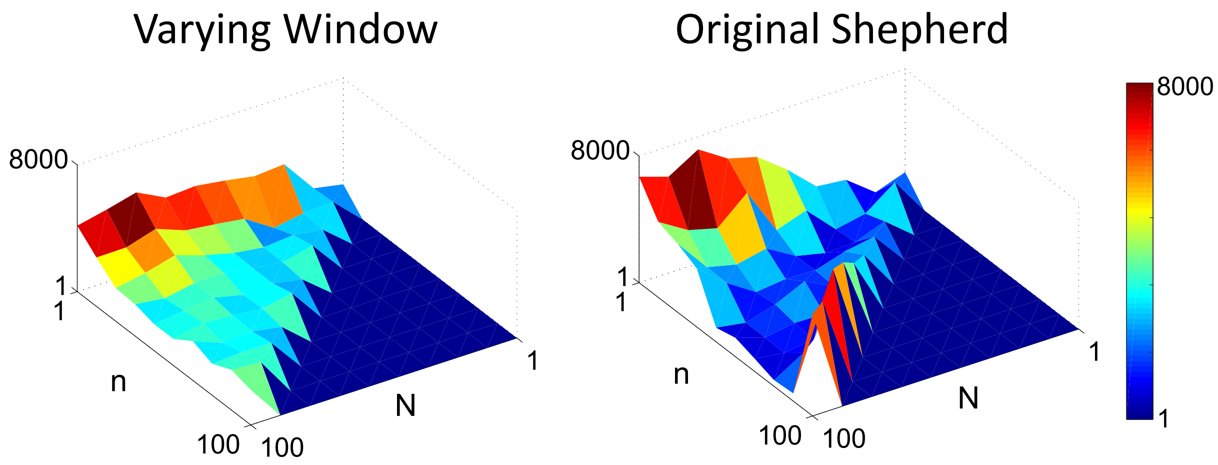


Figure 4.9: Time to target (without failed trials) for a number of agents (N) vs a number of neighbours (n). Left panels shows this time for the SPRT+MSPRT shepherd and right panel for the Strömbom et al. (2014) shepherd

Figure 4.9 provides some insight into why the MSPRT shepherd outperformed the original one. The panel on the left is consistent with those of fig. 4.7, where the decision time was shown to increase as $n << N - 1$. The panel on the right, which corresponds to the original shepherd, shows a significantly higher time throughout the parameter range, as well as a lot of variability, potentially due to the simulations timing out and contributing negatively to the average.

In ethological situations, due to the nature of stimuli, highly complex processes (such as the MSPRT) are not always necessary since some decisions are easy enough to be computed instantly. However, more complicated decisions where accuracy is prioritised can benefit from such a decision mechanism over a simple immediate decision.

In our view, for the shepherd-sheep scenario, the most important of the decisions (where higher accuracy was required) was which sheep to collect. This was due to the fact that collecting a sheep that wasn't the furthest from the centre of mass could lead to the furthest sheep distancing itself even more. This distancing would also impact the behaviour of neighbouring sheep, resulting in undesirable group behaviour. A sequence of wrong choices would then make it more difficult if not impossible for the shepherd to bring the flock to a cohesive group and herd them to the target.

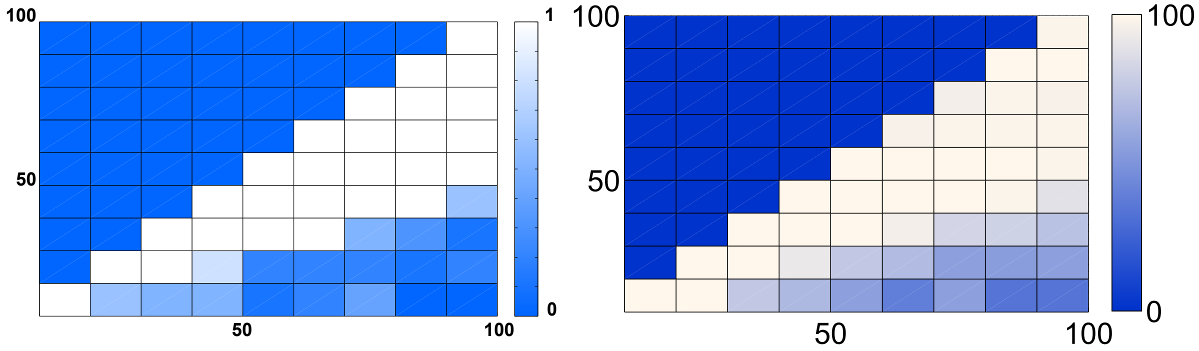


Figure 4.10: Proportion of successful shepherding events for a number of agents (N - x axis) vs a number of neighbours (n - y axis). Left panel shows results with noise free evidence, right panel shows results with noisy evidence

In figure fig. 4.5, the similarities between the original shepherd model and the MSPRT-like implementation in noise-free scenarios were demonstrated, where the two models were shown to display the same pattern of results. Comparing the MSPRT-like implementation in noisy and noise free scenarios, in figure fig. 4.10, one can note that virtually the same pattern of results is displayed; this indicates that the addition of noise to the decision was compensated, as expected, by the use of integration windows, allowing the algorithm to remain functionally the same.

Through the extrapolation of the biologically accurate sheep behaviour model to introduce noise, we have shown that the shepherding algorithm as originally presented is unable to cope with the variability in the inputs, therefore demonstrating it's narrow field of application. When implemented with an MSPRT algorithm as the decision mechanism, we showed that the shepherd was able to cope with larger uncertainty in sheep position without compromising accuracy.

Particular theorised patterns of behaviour, such as the selfish-herd theory, displayed by sheep were shown to have parallels in other species; particularly behaviours which are motivated by survival from predators; this indicates that perhaps certain models which are designed for application to a particular group could potentially be successfully applied to others. Due to the wide range of applications of the MSPRT algorithm and this model's ability to produce satisfactory results with noisy inputs, we believe it could be applied to other groups with distinct behavioural patterns with high success rates.

Chapter 5

Discussion

The algorithm described in this thesis fills a gap in the decision making literature by extrapolating the use of trial based models in continuous decision making without constraints on trial length or input dynamics (Nunes and Gurney, 2016). As a standalone algorithm, the MSPRT requires a mechanistic reset and therefore cannot be used continuously to make successive decisions (Baum and Veeravalli, 1994; Dragalin et al., 1999, 2000; McMillen and Holmes, 2005); our algorithm, however, implemented an integration window mechanism where only the N more recent samples of evidence were used to compute a decision at any point in time, which allowed it to effectively discard outdated evidence, therefore enabling its use in continuous tasks.

When designing our windowing mechanism, we initially tested a number of window sizes and threshold values, with accumulated reward measured at fixed points of the trial (every 1000 ms). As expected, results indicated that overall amount of selection was increased for larger windows with higher thresholds since more evidence was accumulated and therefore the signals were more likely to cross the threshold within the allotted time (Bogacz et al., 2006). Higher thresholds increased the performance of smaller windows since they helped counteract the limited amount of evidence accumulated, albeit also increasing incorrect selection.

Due to the varying nature of the stimuli, we found that no fixed window was ideal across all tasks since the competition was stronger in some than it was in others and therefore the decisions could be made easier or harder depending on the structure of the inputs. After identifying a relationship between the decision time and the difficulty of decision (defined by us as the difference in means between the competing inputs), and modelling this using a power law, we hypothesized that this relationship could be used as a principled way of setting the window size, conferring the algorithm an extra level of abstraction from human input since it would then become capable of adapting to the decision at hand without the need for a parameter reset.

Results indicated that a problem dependent window allowed the algorithm to maintain high levels of correct selection and that the use of the difficulty of decision as a setting parameter for the rate at which evidence was discarded presented a suitable solution to a problem that affects the vast majority of decision making algorithms proposed in the literature: the lack of evidence refreshing mechanisms; since the window size was varied at run time with the variation of the inputs to the algorithm, an excess of input samples had to be retained at all times (up to a fixed maximum) since in order for a change in window size to take effect immediately at any point in the trial, the number of samples required had to be available.

One can note that for extremely difficult decisions, the costs in terms of memory requirements would be significant since a very large number of samples would be required by default. In order to overcome this, we used the window generating mechanism to generate a decay parameter which worked similarly to leak parameters in the literature, where only the sum of all the evidence was required; the evidence would then be decayed over time using the dynamically generated parameter; this resulted in lower levels of reward, but was a much simpler and memory efficient solution.

After establishing the usability of the windowing mechanism within the rigorous implementation of the MSPRT, we explored different input scenarios which did not obey the strict derivation of this algorithm, but rather constituted an empirical study of the boundaries of the applications of strict theoretical Sequential Probability Ratio Tests (Wald and Wolfowitz, 1948; Ratcliff et al., 1978; Smith and Vickers, 1988; Schall, 2001; Ratcliff et al., 1999).

Since the gain parameter of the algorithm was derived from the mean value of the competing inputs, we started by analysing the effect of changing this in the overall quality of results. Results indicated that even with the inputs being allowed to vary their means over time, the decision times were maintained, suggesting that the algorithm was capable of handling the extra variability (results were compared to those obtained by Bogacz and Gurney (2007)).

Building on the results from the variation of the gain and given Lepora and Gurney (2012)'s finds on the use of arbitrary alternatives with the MSPRT algorithm, we devised subsequent empirical studies to look at other ways in which we could approximate our highly constrained inputs to those in a realistic decision scenarios. For this, we allowed the inputs to vary outside of the winner-loser pattern required for a strict MSPRT implementation where the inputs are characterised by a winning channel with a high mean and N losing channels with the same low mean.

Interestingly, breaking the theoretical constraints by increasing input variability led to improved performance through a decrease in decision time.

In order to bridge the abstract laboratory experiments where the inputs tend to be defined as 'signals' with the real decisions performed by animals, we took an existing model of shepherd-sheep behaviour, developed by Strömbom et al. (2014) and embedded our full MSPRT in place of the authors' instant decision mechanism; this allowed the shepherd to perform decisions based on evidence accumulated over time, thus removing the constraint of perfect information being made available to it.

One of the main advantages of combining the MSPRT inspired algorithm presented previously with the Strömbom et al. (2014) model was that this allowed for our algorithm to be tested in a more challenging scenario than those explored up to this point; since the original model was designed to fit observed patterns of behaviour, the sheep display complex behavioural patterns of repulsion and attraction as well as random movements, which in itself increased the difficulty of the decisions presented to the mechanism.

At any point in time the shepherd had to decide between collecting or herding as well as which sheep to target for each action (collecting particular sheep or positioning itself at a particular position given the location of all agents); a shepherding event was considered successful when the centre of mass of the sheep was placed in the target area within 8000 time steps.

Different window sizes and geometries were tested, with the relationship between speed and accuracy, and time to decision used as a term of comparison between implementations. Windows of rectangular geometry displayed a proportional increase in decision time with the increase of trial length; conversely, exponential windows displayed an inversely proportional relationship between decision time and length of trial, indicating that longer evidence accumulation times improve accuracy of decision and therefore result in quicker trials.

In line with the results in previous chapters, the exponential window with $\lambda = 0.01$ yielded the optimal balance between speed and accuracy.

Initial experiments comparing our implementation of the shepherd (using an MSPRT-like decision mechanism) to the original shepherding algorithm yielded the same pattern of results, indicating a parallel between the two; in a later stage, an implementation of the algorithm with noisy inputs and an evidence discarding mechanism proved a near perfect approximation of the initial results; this could be interpreted as a more realistic approximation of the shepherd behaviour without the assumption of perfect knowledge.

The algorithm described in our work makes use of evidence accumulated with no notion of context; in this case, the overall value of the alternatives is never taken into account as only the difference between them is used to compute the decision. Our studies indicated that, in certain simulations, particularly when the number of neighbours is significantly smaller than the total number of agent, deadlock situations can arise, where the decision maker (shepherd) becomes unable to decide how to proceed. In situations where the difference between competing options is negligible, to the point of causing a deadlock, it would make sense for the decision maker to choose randomly, favouring decision time over accuracy, This could be achieved through the addition of random movement of the agent, which would trigger actions in the sheep and therefore break the current deadlock.

The simplistic nature of the mechanism described previously does not take into the account the possibility of all competing options having low magnitude, in which case the shepherd could potentially benefit from holding off the decision until a new, better option is available. The behaviour described previously is known as value-based, or value sensitive decision making.

Value based decision making has been widely studied in the last decade; not only in how it relates to computational decision making (Bogacz et al., 2007; Milosavljevic et al., 2010; Krajbich and Rangel, 2011; Tsetsos et al., 2012) but also to human neurobiology (Rangel et al., 2008; Hikosaka, 2010; Wallis, 2011; Vlaev et al., 2011). The LCA model described by Bogacz et al. (2007), which approximates the SPRT when inhibition and decay are high, uses a non-linear utility function based on Bernoulli's logarithmic nonlinearity to model complex human risk aversion behaviours and Weber's law to equip the decision making algorithm with a value account of competing options. Despite the MSPRT's lack of sensitivity to value, future work could be done to adapt the mechanism in a similar way to that previously described, in which value is given to the inputs based on a number of environmental and intrinsic factors to the mechanism.

In Simen et al. (2009), authors argued that the optimality of the threshold chosen by Sequential Probability Ratio Test might be lost for situations which involve non trial based setting where many decisions are made in sequence; indicating that a varying threshold could be used to keep error rates down whilst maintaining the lowest possible decision time.

Time-varying threshold were discussed in the literature with regards to many computational models and real observations (Malhotra et al., 2017); one of these models, which was discussed in the Background chapter of this thesis is the Diffusion Model (Ditterich, 2006; Simen et al., 2009). In Ditterich (2006), the authors discussed decision models which allow for decisions to me performed with less evidence accumulated; a number of models were presented, with one of them (model 2) being equipped with a time-varying threshold. The mechanism described by the authors lowered the decision threshold as the trial advanced, effectively decreasing the amount of evidence needed to make a decision; this was able to account for behavioural but not psychological data. Such a mechanism could be utilised by our algorithm to elicit decisions both in situations where the competing options are closely matched and situations where there is very little evidence.

Despite the extensive computational decision making literature and the claims on parallels with human decision making, most algorithms are applied to extremely constrained scenarios that bare little to no parallel to day to day decisions performed by humans and animals; as such there still is a clear need for the exploration of algorithms with real applications to less constrained scenarios. Our empirical studies of and MSPRT-like algorithm implemented in less contrived settings have demonstrated its versatility in a wide range of problems, indicating that this could be well suited for more real world problems that don't require human setup or a human observer; it also provides evidence to show that despite some algorithms being designed under extremely limiting circumstances, there are a multitude of ways in which these can be extended to work in more generic conditions.

However realistic the model of sheep behaviour is, it is arguably only representative of the species observed by the authors. In follow up experiments, different groups of agents could be identified which displayed different behavioural patterns; these could be groups of birds or fish (Gade, Shripad; Paranjape, Aditya; Chung, 2015; Strömbom et al., 2015), usually modelled by self-propelled particle models. More complex behaviours such as Kin selection (Jane and Eberhard, 1975) identified among different groups of vertebrates and invertebrates, could also pose interesting challenges since they model self-sacrifice (or altruism), a more complex behavioural pattern which could have a parallel with human group behaviour. If successfully applied, the algorithm would then constitute a solid model of interaction with generic groups and could potentially be applied in the guidance of humans.

In summary, even though the MSPRT is considered a trial based algorithm, and has mostly been applied to limited, trial based scenarios, it displays a lot of potential to be used continuously for real world applications. By initially testing the model in the theoretical framework presented and secondly breaking the theoretical constrains, we have designed an algorithm capable of allowing a biologically inspired shepherd model, previously implemented only in noise-free scenarios, to withstand noise with negligible costs to performance, despite the loss of theoretical rigour of a strict MSPRT algorithm. We believe the model encompasses the best characteristics of the MSPRT whilst still being representative of a real life decision making algorithm.

Bibliography

Aggarwal, A. and Goatin, P. (2016). Crowd dynamics through non-local conservation laws. *Bulletin of the Brazilian Mathematical Society, New Series*, 47(1):37–50.

Arnold, G. and Pahl, P. (1974). Some aspects of social behaviour in domestic sheep. *Animal Behaviour*, 22(3):592–600.

Arnold, G., Wallace, S., and Rea, W. (1981). Associations between individuals and home-range behaviour in natural flocks of three breeds of domestic sheep. *Applied Animal Ethology*, 7(3):239–257.

Baum, C. and Veeravalli, V. (1994). A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory*, 40(6):1994–2007.

Bennett, B. and Trafankowski, M. (2012). A comparative investigation of herding algorithms. In Antony Galton and Zena Wood, editor, *UNDERSTANDING AND MODELLING COLLECTIVE PHENOMENA*, pages 33–38, Birmingham.

Bitzer, S., Park, H., Blankenburg, F., and Kiebel, S. J. (2014). Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model. *Frontiers in human neuroscience*, 8.

Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. *Trends in Cognitive Sciences*, 11(3):118.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, 113(4):700–65.

Bogacz, R. and Gurney, K. (2007). The Basal Ganglia and Cortex Implement Optimal Decision Between Alternative Actions. *Neural Computation*, 477(2):442–477.

Bogacz, R., Usher, M., Zhang, J., and McClelland, J. L. (2007). Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 362(1485):1655–1670.

Bouchard, M., Haegele, J., and Hexmoor, H. (2015). Crowd dynamics of behavioural intention: train station and museum case studies. *Connection Science*, 27(2):164–187.

Brulé, J., Engel, K., Fung, N., and Julien, I. (2016). Evolving Shepherding Behavior with Genetic Programming Algorithms.

Busemeyer, J. and Townsend, J. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, 100(3):432.

Caballero, J. A., Lepora, N. F., and Gurney, K. N. (2015). Probabilistic Decision Making with Spikes: From ISI Distributions to Behaviour via Information Gain. *PLOS ONE*, 10(4):e0124787.

Chevalier, G. and Deniau, J. M. (1990). Disinhibition as a basic process in the expression of striatal functions. *Trends in neurosciences*, 13(7):277–80.

Chevalier, G., Vacher, S., Deniau, J. M., and Desban, M. (1985). Disinhibition as a basic process in the expression of striatal functions. I. The striato-nigral influence on tecto-spinal/tecto-diencephalic neurons. *Brain research*, 334(2):215–26.

Cho, H., Jö Nsson, H., Campbell, K., Melke, P., Williams, J. W., Jedynak, B., Stevens, A. M., Groisman, A., and Levchenko, A. (2007). Self-Organization in High-Density Bacterial Colonies: Efficient Crowd Control. *PLoS Biology*, 5(11):2614–2623.

Cho, R. Y., Nystrom, L. E., Brown, E. T., Jones, A. D., Braver, T. S., Holmes, P. J., and Cohen, J. D. (2002). Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. *Cognitive, affective & behavioral neuroscience*, 2(4):283–99.

Churchland, A. K. and Ditterich, J. (2012). New advances in understanding decisions among multiple alternatives. *Current opinion in neurobiology*, 22(6):920–926.

Churchland, A. K., Kiani, R., and Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature neuroscience*, 11(6):693–702.

Cowling, P. and Gmeinwieser, C. (2010). AI for Herding Sheep. *Association for the Advancement of Artificial Intelligence*.

Cristiani, E., Piccoli, B., and Tosin, A. (2011). Multiscale Modeling of Granular Flows with Application to Crowd Dynamics. *Multiscale Modeling & Simulation*, 9(1):155–182.

Ding, L. and Gold, J. (2013). The Basal Ganglia's Contributions to Perceptual Decision Making. *Neuron*, 79(4):640–649.

Ditterich, J. (2006). Evidence for time-variant decision making. *The European journal of neuroscience*, 24(12):3628–41.

Ditterich, J. (2010). A Comparison between Mechanisms of Multi-Alternative Perceptual

Decision Making: Ability to Explain Human Behavior, Predictions for Neurophysiology, and Relationship with Decision Theory. *Frontiers in neuroscience*, 4:184.

Dragalin, V. P., Tartakovsky, A. G., Veeravalli, V. V., and Member, S. (1999). Multihypothesis Sequential Probability Ratio Tests Part I : Asymptotic Optimality. *IEEE Transactions on Information Theory*, 45(7):2448–2461.

Dragalin, V. P. V., Tartakovsky, A. A. G., Veeravalli, V. V., and Member, S. (2000). Multihypothesis Sequential Probability Ratio Tests Part II : Accurate Asymptotic Expansions for the Expected Sample Size. *IEEE Transactions on Information Theory*, 46(4):1366–1383.

Evered, M., Burling, P., and Trotter, M. (2014). An Investigation of Predator Response in Robotic Herding of Sheep. In *International Conference on Intelligent Agriculture*, pages 49–54, Singapore.

Feldman, J. A. and Ballard, D. H. (1982). Connectonist Models and Their Properties. 254:205–254.

Gade, S., Paranjape, A. A., and Chung, S.-J. (2016). Robotic Herding Using Wavefront Algorithm: Performance and Stability. In *AIAA Guidance, Navigation, and Control Conference*, Reston, Virginia. American Institute of Aeronautics and Astronautics.

Gade, Shripad; Paranjape, Aditya; Chung, S.-J. (2015). Herding a Flock of Birds Approaching an Airport Using an Unmanned Aerial Vehicle. In *AIAA SciTech, Guidance Navigation and Control Conference*, number January, pages 1–17.

Gold, J. I. and Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in cognitive sciences*, 5(1):10–16.

Gold, J. I. and Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, 30:535–74.

Gurney, K., Prescott, T. J., and Redgrave, P. (2001a). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological cybernetics*, 84(6):401–10.

Gurney, K., Prescott, T. J., and Redgrave, P. (2001b). A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. *Biological cybernetics*, 84(6):411–23.

Heekeren, H. R., Marrett, S., and Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature reviews. Neuroscience*, 9(6):467–79.

Helbing, D., Buzna, L., Johansson, A., and Werner, T. (2005). Self-Organized Pedestrian Crowd Dynamics: Experiments, Simulations, and Design Solutions. *Transportation Science*, 39(1):1–24.

Helbing, D., Molnár, P., Farkas, I. J., and Bolay, K. (2001). Self-organizing pedestrian movement. *Environment and Planning B: Planning and Design*, 28:361–383.

Hikosaka, O. (2010). The habenula: from stress evasion to value-based decision-making. *Nature Reviews Neuroscience*, 11(7):503–513.

Hunter, R. and Milner, C. (1963). The behaviour of individual, related and groups of South Country Cheviot hill sheep. *Animal Behaviour*, 11(4):507–513.

Jane, M. and Eberhard, W. (1975). The Evolution of Social Behavior by Kin Selection. *The Quarterly Review of Biology*, 50(1):1–33.

Jyh-Ming Lien, Rodriguez, S., Malric, J., and Amato, N. (2005). Shepherding Behaviors with Multiple Shepherds. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 3402–3407. IEEE.

Kennerley, S. W., Walton, M. E., Behrens, T. E. J., Buckley, M. J., and Rushworth, M. F. S. (2006). Optimal decision making and the anterior cingulate cortex. *Nature neuroscience*, 9(7):940–7.

King, A. J., Wilson, A. M., Haddadi, H., Hailes, S., and Morton, A. J. (2012). Selfish-herd behaviour of sheep under threat. *Current Biology*, 22(14).

Krajbich, I. and Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33):13852–7.

Krause, J. and Ruxton, G. (2002). *Living in groups.*

Lakshika, E., Barlow, M., and Easton, A. (2015). Evolving High Fidelity Low Complexity Sheepdog Herding Simulations Using a Machine Learner Fitness Function Surrogate for Human Judgement. pages 330–342. Springer International Publishing.

Laming, D. (1968). *Information theory of choice-reaction times.*

Langston, P. A., Masling, R., and Asmar, B. N. (2006). Crowd dynamics discrete element multi-circle model. *Safety Science*, 44(5):395–417.

Lepora, N. F., Fox, C. W., Evans, M. H., Diamond, M. E., Gurney, K., and Prescott, T. J. (2012). Optimal decision-making in mammals: insights from a robot study of rodent texture discrimination. *Journal of the Royal Society, Interface / the Royal Society*, 9(72):1517–28.

Lepora, N. F. and Gurney, K. N. (2012). The Basal Ganglia Optimize Decision Making over General Perceptual Hypotheses. *Neural Computation*, 24(11):2924–2945.

Malhotra, G., Leslie, D. S., Ludwig, C. J. H., and Bogacz, R. (2017). Overcoming indecision by changing the decision boundary. *Journal of Experimental Psychology: General*, 146(6):776–805.

McMillen, T. and Holmes, P. (2005). The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, 50(1):30–57.

Mehner, W., Boltes, M., Mathias, M., and Leibe, B. (2015). Robust Marker-Based Tracking for Measuring Crowd Dynamics. pages 445–455. Springer International Publishing.

Miki, T. and Nakamura, T. (2006). An Effective Simple Shepherding Algorithm Suitable for Implementation to a Multi-Mmobile Robot System. In *First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC'06)*, volume 3, pages 161–165. IEEE.

Miki, T. and Nakamura, T. (2007). An effective rule-based shepherding algorithm by using reactive forces between individals. *International Journal of Innovative Computing*, 3(4):813–823.

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., and Rangel, A. (2010). The Drift Diffusion Model Can Account for the Accuracy and Reaction Time of Value-Based Choices Under High and Low Time Pressure. *SSRN Electronic Journal*.

Moussaïd, M., Helbing, D., Garnier, S., Johansson, A., Combe, M., and Theraulaz, G. (2009). Experimental study of the behavioural mechanisms underlying self-organization in human crowds. In *Proceedings. Biological sciences / The Royal Society*, volume 276, pages 2755–62. The Royal Society.

Nunes, L. F. and Gurney, K. (2016). Multi-alternative decision-making with non-stationary inputs. *Royal Society Open Science*, 3(8):160376.

Ossmy, O., Moran, R., Pfeffer, T., Tsetsos, K., Usher, M., and Donner, T. H. (2013). The timescale of perceptual evidence integration can be adapted to the environment. *Current biology : CB*, 23(11):981–6.

Pirrone, A., Stafford, T., and Marshall, J. A. R. (2014). When natural selection should optimize speed-accuracy trade-offs. *Frontiers in Neuroscience*, 8(73):73.

Prescott, T. J., Montes González, F. M., Gurney, K., Humphries, M. D., and Redgrave, P. (2006). A robot model of the basal ganglia: behavior and intrinsic processing. *Neural networks : the official journal of the International Neural Network Society*, 19(1):31–61.

Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9(7):545–556.

Ratcliff, R., Cherian, A., and Segraves, M. (2003). A Comparison of Macaque Behavior and Superior Colliculus Neuronal Activity to Predictions From Models of Two-Choice Decisions. *Journal of Neurophysiology*, 90(3):1392–1407.

Ratcliff, R. and McKoon, G. (2008). The Diffusion Decision Model : Theory and Data for Two-Choice Decision Tasks. 922:873–922.

Ratcliff, R. and Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, 9(5):347–356.

Ratcliff, R. and Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1):127–140.

Ratcliff, R., Schachter, S., and Singer, J. (1978). A Theory of Memory Retrieval. *Psychological review*, (2).

Ratcliff, R. and Smith, Philip, L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological Review*, 111(2):333–367.

Ratcliff, R., Van Zandt, T., and McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological review*, 106(2):261–300.

Redgrave, P., Prescott, T. J., and Gurney, K. (1999). The Basal Ganglia: A Vertebrate Solution to the Selection Problem? 89(4):1009–1023.

Reina, A., Marshall, J. A. R., Trianni, V., and Bose, T. (2016). Model of the best-of-N nest-site selection process in honeybees. *Physical Review E*, 95(5):052411.

Resulaj, A., Kiani, R., Wolpert, D. M., and Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, 461(7261):263–266.

Roe, R. M., Busemeyer, J. R., and Townsend, J. T. (2001). Multialternative Decision Field Theory: A Dynamic Connectionist Model of Decision Making. *Psychological review*, 108(2):370–392.

Schall, J. D. (2001). Neural correlates of choosing. *Nature Reviews Neuroscience*, 34(2).

Shadlen, M. N., Newsome, W. T., Wong, A. L., Haith, A. M., Krakauer, J. W., Werklebergner, M., Grandy, T. H., Chicherio, C., and Schmiedek, F. (2001). Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey. pages 1916–1936.

Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., and Cohen, J. D. (2009). Reward rate optimization in two-alternative decision making: Empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and Performance*, 35(6):1865–1897.

Singh, H., Arter, R., Dodd, L., Langston, P., Lester, E., and Drury, J. (2009). Modelling subgroup behaviour in crowd dynamics DEM simulation. *Applied Mathematical Modelling*, 33(12):4408–4423.

Smith, P. L. and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in neurosciences*, 27(3):161–8.

Smith, P. L. and Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, 32(2):135–168.

Strömbom, D., Mann, R. P., Wilson, A. M., Hailes, S., Morton, A. J., Sumpter, D. J. T., King, A. J., Tinbergen, N., Couzin, I., Krause, J., James, R., Ruxton, G., Franks, N., Sumpter, D., Aoki, I., Hamilton, W., King, A., Wilson, A., Wilshin, S., Lowe, J., Haddadi, H., Hailes, S., Morton, A., Lien, J., Bayazit, O., Sowell, R., Rodriguez, S., Amato, A., Lien, J., Rodriguez, S., Malric, J., Amato, N., Barry, A., Dalrymple-Smith, H., Adachi, Y., Kakikura, M., Miki, T., Nakamura, T., Vo, C., Harrison, J., Lien, J., Kachroo, P., Vaughan, R., Sumpter, N., Frost, A., Cameron, S., Vaughan, R., Sumpter, N., Henderson, J., Frost, A., Cameron, S., Bennett, B., Trafankowski, M., Hughes, R., Hughes, R., Fingas, M., Schultz, A., Adams, W., Trust, E. W., Turgut, A., Celikkanat, H., Gökce, F., Sahin, E., Reynolds, C., Huth, A., Wissel, C., Lukeman, R., Li, Y., Edelstein-Keshet, L., Buhl, J., Sumpter, D., Couzin, I., Hale, J., Despland, E., Miller, E., Simpson, S., Vicsek, T., Zafeiris, A., Lien, J., Pratt, E., Coppinger, L., Coppinger, R., Darling, F., Haddadi, H., King, A., Willis, A., Fay, D., Lowe, J., Morton, A., Hailes, S., Wilson, A., Strömbom, D., Moussaid, M., Helbing, D., Theraulaz, G., Ballerini, M., Zahugi, E., Shanta, M., Prasad, T., Isobe, M., Helbing, D., Nagatani, T., Procaccini, A., Bailey, I., Myatt, J., and Wilson, A. (2014). Solving the shepherding problem: heuristics for herding autonomous, interacting agents. *Journal of the Royal Society, Interface / the Royal Society*, 11(100):20140719.

Strömbom, D., Siljestam, M., Park, J., and Sumpter, D. (2015). The shape and dynamics of local attraction. *The European Physical Journal Special Topics*, 224(17-18):3311–3323.

Tsetsos, K., Gao, J., Mcclelland, J. L., Usher, M., Lagnado, D. A., Huk, A. C., and Johnson, J. G. (2012). Using Time-Varying Evidence to Test Models of Decision Dynamics: Bounded Diffusion vs. the Leaky Competing Accumulator Model. *Frontiers in neuroscience*, 6(June):79.

Tsetsos, K., Usher, M., and Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological review*, 117(4):1275–93.

Tsetsos, K., Usher, M., and McClelland, J. L. (2011). Testing multi-alternative decision models with non-stationary evidence. *Frontiers in neuroscience*, 5(May):63.

Usher, M. and McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3):550–592.

Usher, M., Tsetsos, K., Yu, E. C., and Lagnado, D. a. (2013). Dynamics of decision-making: from evidence accumulation to preference and belief. *Frontiers in psychology*, 4(October):758.

van Maanen, L., Grasman, R. P. P. P., Forstmann, B. U., and Wagenmakers, E.-J. (2012). Piéron's Law and Optimal Behavior in Perceptual Decision-Making. *Frontiers in neuroscience*, 5(January):143.

van Ravenzwaaij, D., van der Maas, H. L. J., and Wagenmakers, E.-J. (2012). Optimal decision making in neural inhibition models. *Psychological review*, 119(1):201–15.

Vaughan, R., Sumpter, N., Henderson, J., Frost, A., and Cameron, S. (2000). Experiments in automatic flock control. *Robotics and Autonomous Systems*, 31(1):109–117.

Veliz-Cuba, A., Kilpatrick, Z. P., Josi, K., Josić, K., and Alan Veliz-Cuba Zachary P. Kilpatrick, K. J. (2015). Stochastic Models of Evidence Accumulation in Changing Environments. *SIAM Review*, 58(2):264–289.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, 13:37–58.

Vlaev, I., Chater, N., Stewart, N., and Brown, G. D. A. (2011). Does the brain calculate value? *Trends in Cognitive Sciences*, 15:546–554.

Wald, A. and Wolfowitz, J. (1948). Optimum Character of the Sequential Probability Ratio Test. *The Annals of Mathematical Statistics*, 19(3):326–339.

Wallis, J. D. (2011). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature Neuroscience*, 15(1).

Wiesenfeld, K. and Moss, F. (1995). Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs. *Nature*, 373(6509):33–36.