JMLR: Workshop and Conference Proceedings vol 26 (2012) 37–61 On-line Trading of Exploration and Exploitation 2

Exploration and Exploitation with Insufficient Resources

Chris Lovell CJL3@ECS.SOTON.AC.UK
Gareth Jones GJ07R@ECS.SOTON.AC.UK
Klaus-Peter Zauner KPZ@ECS.SOTON.AC.UK
Steve R. Gunn SRG@ECS.SOTON.AC.UK
School of Electronics and Computer Science,
University of Southampton, UK

Editor: Dorota Głowacka, Louis Dorard and John Shawe-Taylor

Abstract

In physical experimentation, the resources available to discover new knowledge are typically extremely small in comparison to the size and dimensionality of the parameter spaces that can be searched. Additionally, due to the nature of physical experimentation, experimental errors will occur, particularly in biochemical experimentation where the reactants may undetectably denature, or reactant contamination could occur or equipment failure. These errors mean that not all experimental measurements and observations will be accurate or representative of the system being investigated. As the validity of observations is not guaranteed, resources must be split between exploration to discover new knowledge and exploitation to test the validity of the new knowledge. Currently we are investigating the automation of discovery in physical experimentation, with the aim of producing a fully autonomous closed-loop robotic machine capable of autonomous experimentation. This machine will build and evaluate hypotheses, determine experiments to perform and then perform them on an automated lab-on-chip experimentation platform for biochemical response characterisation. In the present work we examine how the trade-off between exploration and exploitation can occur in a situation where the number of experiments that can be performed is extremely small and where the observations returned are sometimes erroneous or unrepresentative of the behaviour being examined. To manage this trade-off we consider the use of a Bayesian notion of surprise, which is used to perform exploration experiments whilst observations are unsurprising from the predictions that can be made and exploits when observations are surprising as they do not match the predicted response.

Keywords: Limited resources, exploration–exploitation, Bayesian surprise

1. Introduction

In physical experimentation, the resources typically available are generally small in comparison to the size and scale of the parameter space. For example there may perhaps be only a handful of experiments available per parameter dimension. In general the amount of resources can be considered as being 'not enough' to provide a highly confident prediction of the behaviour being observed. Therefore the goal is to get a good reliable prediction of the observable behaviours, with as few experiments as possible. To aid the experimenter, statistical machine learning techniques can be employed to perform pattern analysis on the data available and choose the experiments to perform, with the goal of maximising the information gained whilst minimising the resources spent. These techniques are similar

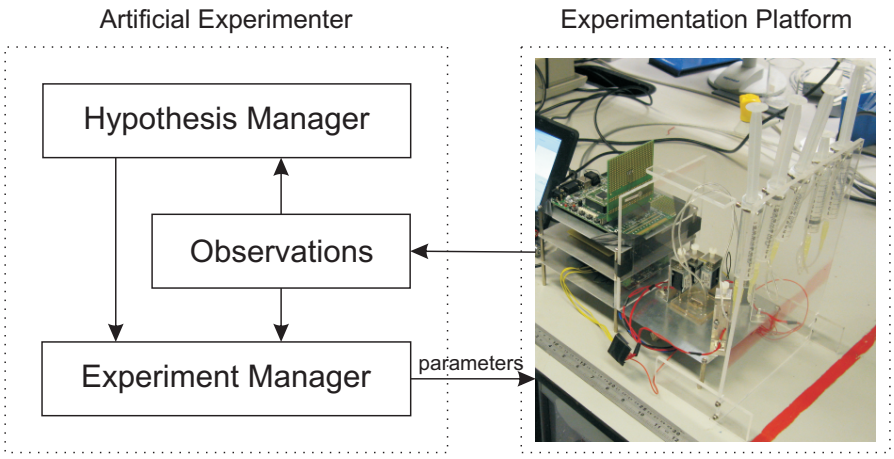

Figure 1: Overview of an artificial experimenter combined with an automated experimentation platform to allow for autonomous experimentation. The artificial experimenter uses the information available to build hypotheses and determine experiments to perform. It provides experiment parameters to perform and is provided the observational results for those experiments. Also shown is a fully automated lab-on-chip based experimentation platform under development. The syringes on the right of the device hold the liquid reactants available. The flow of reactants is controlled by on-chip valves driven by computer controlled solenoids. The platform contains a UV photometer that allows for measurements of optical absorbance to be taken for reactants flowing within the microfluidic chip. The measurements are the observations returned.

to computational scientific discovery (Langley et al., 1987) and active learning (MacKay, 1992; Cohn et al., 1994; Settles, 2009). We label the combination of techniques that are implemented to perform this pattern recognition and adaptive experiment selection within a laboratory problem, as an artificial experimenter (Lovell et al., 2010). When combined with automated hardware capable of performing the experiments requested, an autonomous experimentation machine can be created as illustrated in Figure 1.

Additionally to the limited resources problem, physical experimentation has the problem that experimental errors or unexpected undetectable physical changes in the reactants can occur, which can yield observations not representative of the behaviour being investigated. These erroneous observations can be thought of as being outliers, except that there will be insufficient data available to identify them as such with any degree of confidence. One approach to handling this uncertainty in the observation validity, is to consider multiple hypotheses in parallel that have different views about the data (Lovell et al., 2010). The information within these hypotheses can be exploited to select experiments where they most disagree, in order to obtain experiments that can dismiss invalid hypotheses. However experiments must also be performed that can search for features of the behaviour that have not yet been characterised. The artificial experimenter has to make a continual trade-off

between performing experiments to discover features of the behaviour not yet identified, with spending resources to ensure that the current models of the behaviour being investigated are accurate. With too little exploration, features of the behaviour such as local maxima and minima will be missed, resulting in poor scope of the hypotheses. Whilst with too little exploitation, inaccuracies and errors will occur in the hypotheses caused by the erroneous observations, resulting in hypotheses with poor accuracy.

We argue that in a situation where the resources are extremely limited and the observations may not always be representative of the true underlying behaviour, an experiment selection strategy will want to disprove invalid hypotheses when there is disagreement about the behaviour or validity of an observation and search for new discoveries when they agree. In other words, exploit the information held within the hypotheses to select experiments that maximise the discrimination between the hypotheses when there are good hypotheses supported by the experimental evidence that disagree. Whilst exploring the parameter space when there is no disagreement between good hypotheses. To manage this trade-off we use a Bayesian formulation of surprise, first used to identify surprising occurrences in video sequences (Itti and Baldi, 2009). In this work, the surprise is used such that exploitation occurs when the previous experiment was surprising and exploration occurs when the previous experiment was not surprising. The use of surprise to manage this trade-off is analogous to techniques performed by successful human scientists, who will perform focussed experiments to learn why an experiment yielded a surprising or unexpected observation (Kulkarni and Simon, 1990).

Here we discuss how the trade-off between exploration and exploitation has been considered within an artificial experimenter that can perform automatic response characterisation with a small, noisy and potentially erroneous set of observations. In particular enzymatic responses has been considered as the domain to evaluate the techniques, however the design of the artificial experimenter is domain independent. As current active learning techniques do not fully capture the problems and uncertainties faced in physical experimentation, we take the approach of addressing the problem and filling the gaps through attempting to capture how successful scientists operate and make decisions, which make the techniques for managing the trade-off presented here driven more from practise than theory. In Section 2 we briefly re-pose the problem within a multi-armed bandit framework. In Section 3 we briefly discuss how hypotheses are represented within the system. In Section 4 we discuss different methods that have been used for managing the exploration–exploitation trade-off within our artificial experimenter. These techniques are evaluated through a 1-dimensional simulated trial and laboratory trial in Section 5, along with a 2-dimensional simulated trial in Section 6. In Section 7 we discuss how previous systems have addressed the exploration–exploitation trade-off.

2. Description of the Problem within a Multi-Armed Bandit Framework

For ease of understanding the present problem within the exploration-exploitation community, we abstract the problem to one within a multi-armed bandit framework. In the multi-armed bandit problem there are a number of levers, which correspond to different possible experiments, which when pulled return some reward. The reward obtained in experimentation is not directly a monetary reward, but rather an information reward. As-

suming we have multiple hypotheses in consideration and that we take the philosophy of science view that information is obtained by disproving hypotheses (Chamberlin, 1890), an example of a rewarding experiment would be one that will provide information to discriminate between a number of good hypotheses. A prediction of the minimum expected reward can be made by examining the current hypotheses under consideration and determining the difference in predictions for them across the possible experiment parameters. However the actual reward may be much higher than predicted in regions where few experiments have been performed and where the predictions of the hypotheses are not representative of the true behaviour being investigated. Selecting where the hypotheses are maximally different will therefore be equivalent to exploitation. The goal therefore for experimentation is the same as a multi-armed bandit problem, to maximise the reward, or information obtained, over a number of trials or experiments. However not all experiments will produce a reward, take for example the case where all hypotheses agree with the observation obtained and no new information is learnt. Additionally the number of experiments that can be afforded will be many times smaller than the total number of unique experiment parameters or levers. Furthermore the reward available on a particular experiment will generally reduce over time, as the hypotheses under consideration become more accurate in predicting the outcome of an experiment. Although there will be cases where the reward available on a particular experiment may increase over the previous experiment, for example where an erroneous observation occurs that seemingly provides information to disprove a large number of good hypotheses. In cases where a large information gain has occurred, repeat experiments may be useful to determine the validity of an observation, however multiple repeats will lead to a reward tending towards zero. Therefore it is clear that the reward actually obtained by pulling a lever or performing an experiment, would not be modelled by a single static normal distribution within a multi-armed bandit abstraction, but by a distribution where the mean alters over time and there are perhaps two variances, one small variance that provides experimental noise on all experiments, with a second larger variance that provides erroneous observations to some experiments. Due to the differing rewards available over time, the regret function will be similar to that used by Auer (2002), to compare the maximum reward available at time t with the actual reward obtained:

$$\rho = \sum_{t=1}^{T} \left(\max_{x \in \mathcal{X}} \{I_t(x)\} - \hat{i}_t \right) \tag{1}$$

where $\max_{x \in \mathcal{X}} \{I_t(x)\}$ is the maximum reward possible at time t, $\mathcal{X}$ is the set of possible experiments, and $\hat{i}_t$ is the actual reward at time t. Although due to the resource limitation an upper confidence bound approach will not be suitable as it will be unable to initialise the predicted means.

3. Predicting the Behaviour Under Investigation

The present problem of response characterisation can be clearly addressed through using a regression technique. Here spline based regression techniques form the foundation of the hypothesis representation, although alternate techniques could be substituted. Specifically, the techniques used are the smoothing spline in the 1-dimensional problem and thin plate splines in higher dimensions (Wahba, 1990). However, a single hypothesis or distribution

will perform poorly in situations where there are limited observations, with the validity of those observations not guaranteed (Lovell et al., 2010). The poor performance is due to the single hypothesis having to decide on the validity of observations. With little experimental data available, the technique may not be able to correctly identify an outlying observation that is erroneous, causing it to overfit to the error. Alternatively the opposite could occur, where a parameter selection technique like cross-validation may incorrectly identify an apparent outlying observation and ignore it, even though the observation is representative of the behaviour being investigated and it is the hypothesis that it incorrect. Instead we consider the use of multiple hypotheses, similar to query-by-committee (Seung et al., 1992). The multiple hypotheses are used to provide different views about the data that can be considered in parallel, with decisions made about which was the correct decision when more data becomes available. Additionally the multiple hypotheses can be used later in experiment selection, to provide a method for exploitation through selecting where the hypotheses with the most supporting experimental evidence disagree the most.

To deal with the uncertainty caused by only having a limited number of noisy and potentially erroneous observations, a hypothesis manager is used to consider many hypotheses in parallel (Lovell et al., 2010). Each hypothesis can maintain a different view of the behaviour being investigated, along with different views about the validity of observations. In our design, hypotheses go through a process of refinement in cases where observations do not agree with hypotheses, to develop hypotheses that should be more representative of the true underlying behaviour. An observation and hypothesis are identified as being in disagreement when the observation falls outside of the 95% error bar interval for the prediction of the hypothesis. When refining a hypothesis under these circumstances, the system must take into consideration the problem that it will not know whether the disagreement between observation and hypothesis is because the hypothesis is incorrect, or if the observation is erroneous. The refinement process handles this consideration by creating two new refinements of the original hypothesis that the observation disagreed with. In one refinement the disagreeing observation will be declared to be erroneous and the observation will receive a weight of 0, with all other parameters remaining the same. In the other refinement, the observation will be declared to be valid and the observation will be given a high weight, with all other parameters remaining the same. The zero weight will cause the regression to ignore the observation, whilst a high weight will draw the output of the regression curve closer to the observation. The original and two refined hypotheses are all kept in consideration by the hypothesis manager within a working set of hypotheses.

After each experiment is performed, the hypothesis manager creates a new set of hypotheses with random initial parameters to give different starting views of the behaviour being investigated. These hypotheses are added to the working set of hypotheses that were kept in consideration in previous rounds of experimentation. All hypotheses are then compared to all observations to identify any disagreements, where refinements are made to the hypotheses in cases where there are disagreements. Finally all hypotheses are evaluated against all of the previous observations to determine their confidence and quality. By maintaining a working set, or committee of different hypotheses, decisions about the shape of the response or validity of observations can be postponed until sufficient data is available to reject incorrect assumptions. For computational performance, a number of the worst performing hypotheses can be removed.

4. Experiment Selection

The experiment manager determines the experiments to perform using the information available within the hypotheses along with information about where in the parameter space previous experiments have been performed. In choosing experiments, the experiment manager needs to ensure that the parameter space is explored to allow for the discovery of new features of the behaviour being investigated, whilst also making sure that data is obtained that can differentiate between the different hypotheses under consideration to identify the most likely candidate. We do not consider the case where the experiment manager is aware of how many more experiments are available, instead it must assume that the next experiment may be the last experiment so perform the experiment that it decides most useful next. This assumption is made as in experimentation it may not always be clear how many resources will be allocated to a particular problem and may dependent on the observations made. For example experimentation that is obtaining little new information may be terminated earlier than one that is obtaining a large amount of information. Deciding on stopping criteria is also outside of the current investigation into the exploration-exploitation trade-off within experiment selection. Before we consider the trade-off between exploration and exploitation, we briefly define what we mean by a purely exploration and purely exploitation experiment.

Experiments that explore the parameter space should be placed in regions where there are currently no observations available. Often random strategies are used to perform exploration, however with limited resources this may lead to wasted resources in situations where experiments are performed near previously performed experiments that the hypotheses predict well for. The strategy for exploration is therefore to perform experiments whose parameters are maximally distant from any previously performed experiment:

$$\max E(x) = \min_{p \in X} \|x - p\| \tag{2}$$

where X is the set of previously performed experiments.

Experiments that exploit the information held within the hypotheses are used to evaluate the hypotheses. This exploitation should occur through differentiating between as many hypotheses as possible per experiment. To differentiate between hypotheses, the experiment should be chosen where there is the most disagreement between the predictions of the hypotheses. At first glance it may appear that taking the variance of hypothesis predictions would provide the suitable measure of disagreement. However variance can be made to be artificially large in the presence of a single outlying hypothesis prediction, which can cause experiments to be chosen that only differentiate between the outlying hypothesis and all other hypotheses under consideration (Lovell et al., 2010). Alternatively maximising the expected information gain can be used (MacKay, 1992), although for large numbers of hypotheses the calculation can become inefficient. Instead we use a maximum disagreement measure that places experiments where there are differences between high quality hypotheses that currently agree on the previous observations obtained, defined in (Lovell et al., 2010):

$$D(x) = \sum_{i=1}^{|\mathcal{H}|} \sum_{j=1}^{|\mathcal{H}|} \left(1 - P_{h_i}\left(\hat{h}_j(x)|x\right)\right) Q(h_i, h_j) \tag{3}$$

where $\mathcal{H}$ is the set of working hypotheses under consideration, P_{h_i} is the probability that hypothesis h_i agrees with the prediction of hypothesis h_j for experiment parameter x, defined as:

$$P_{h_i}\left(\hat{h}_j(x)|x\right) = \exp\left(\frac{-\left(\hat{h}_i(x) - \hat{h}_j(x)\right)^2}{2\sigma_i^2}\right) \tag{4}$$

where $\hat{h}$(x) is the prediction of a hypothesis for x, σ_i is the error bar at x for hypothesis h_i. The term $Q(h_i, h_j)$ is the measure of quality and agreement between hypotheses, defined as:

$$Q(h_i, h_j) = C(h_i)C(h_j)A(h_i, h_j) \tag{5}$$

where $C(h_i)$ is the confidence of hypothesis h_i based on the previous N observations, defined as:

$$C(h) = \frac{1}{N}\sum_{n=1}^{N} \exp\left(\frac{-\left(\hat{h}(x_n) - y_n\right)^2}{2\tau^2}\right) \tag{6}$$

with τ kept constant at 1.96. The function $A(h_i, h_j)$ is the agreement between the hypotheses for the previous observations obtained, defined as:

$$A(h_i, h_j) = \frac{1}{N}\sum_{n=1}^{N} \exp\left(\frac{-\left(\hat{h}_i(x_n) - \hat{h}_j(x_n)\right)^2}{2\sigma_i^2}\right) \tag{7}$$

with σ_i again being the error bar value of hypothesis h_i for experiment parameter x. The value of $D(x)$ will be high where there are confident hypotheses, which agree on the current available data, but disagree on the outcome of the proposed experiment. By performing an experiment where $D(x)$ is maximal, evidence should be obtained to identify faults within currently well performing hypotheses that have been identified by other hypotheses. This maximum disagreement measure appears to make similar evaluations about disagreement between hypotheses as maximising the information gain, but provides a more efficient calculation.

Next we consider the trade-off between experiments that explore the parameter space and experiments that exploit the information within the hypotheses. In the present problem there is no a-priori experimental evidence available for hypotheses to be built from. Therefore an initial dataset must be obtained. The initial observations are obtained through performing a set of exploratory experiments, which are equidistant across the parameter space. In all trials described here, there are 5 initial exploratory experiments performed, chosen to allow for more resources to be spent on active experiment selection. In the following we consider two techniques for active experiment selection that manage the exploration–exploitation trade-off in different ways. The surprise technique evaluates how surprising the last experiment obtained was, using the surprise to determine whether the next experiment should be a purely exploration or exploitation experiment. Whilst the discrepancy peaks technique attempts to select experiments that have a combined ability to explore and exploit. In both strategies, the artificial experimenter requires a number of exploratory experiments that can be used to generate an initial set of hypotheses. In the 1-dimensional

case presented here the technique performs 5 initial experiments that are equally spaced across the parameter space.

4.1. Selecting Experiments by the Surprise of the Last Experiment

Several previous artificial experimenter techniques discuss the notion of surprise in scientific discovery, and have employed different formulations of surprise to base their experiment selection strategy (Kulkarni and Simon, 1990; Pfaffmann and Zauner, 2001). These techniques are explored further in the related work in Section 7. Surprising observations are important, as they signify that an outcome occurred that was not expected. It could be that the observation was surprising because it was erroneous, which would require investigation to identify the error and remove it from consideration in the hypotheses. Alternatively an observation could be surprising because the current hypotheses are invalid for the behaviour being investigated. In this instance, further investigation should be made in the region of the parameter space where the surprising observation was found, to allow for more representative hypotheses to be made. Regardless of the cause of a surprising observation, further experiments should be performed when a surprising observation is obtained, to investigate why the observation was surprising.

A Bayesian formulation for surprise exists within the background literature, which uses a Kullback-Leibler divergence to identify surprising differences between prior and posterior hypotheses to identify surprising occurrences in video sequences (Itti and Baldi, 2009):

$$S_o = \int_{\mathcal{H}} P\left(h|D\right) \log \frac{P(h|D)}{P(h)} dH \tag{8}$$

where $P(h|D)$ is the posterior probability for the hypothesis and $P(h)$ is the prior probability. However, for use in an artificial experimenter, this surprise function requires an adjustment. In the current form, the equation identifies surprising improvements to the posterior model and scales the result by the posterior confidence. In an experimental setting, as hypotheses can only be disproved (Chamberlin, 1890), we are more interested in those observations that provide evidence that reduce the confidence in previously good hypotheses. In other words, we are interested in observations that disagree with the hypotheses that are currently viewed as being the most accurate representations of the underlying behaviour under investigation. To make this adjustment, we swap the prior and posterior terms in the function (Lovell et al., 2011). Although it may appear counter intuitive to prefer experiments that weaken the confidence of hypotheses, by identifying the inaccuracies of a hypothesis, the hypothesis will subsequently be refined into a new hypothesis that is more representative of the true underlying behaviour being investigated.

We use this metric to calculate the surprise of the most recently obtained observation. To calculate surprise, we take the confidence of the hypotheses before an experiment is performed to provide the prior probability. The posterior probability is taken as the confidence of the same set of hypotheses directly after the experiment has been performed, where the set of observations to evaluate with will now include the observation obtained in that experiment. This allows the surprise of the most recent experiment across the set of

working hypotheses under consideration to be calculated as:

$$S = \sum_{i}^{|\mathcal{H}|} C(h_i) \log \frac{C(h_i)}{C'(h_i)} \tag{9}$$

From this measure of surprise, the decision about whether to explore or exploit next can be made by exploiting when the last observation is surprising, and exploring when the last observation was not surprising, as defined as:

$$x = \begin{cases} D(x) & \text{if } S > 0 \\ E(x) & \text{otherwise} \end{cases} \tag{10}$$

where $E(x)$ is a method for choosing an exploration experiment, for example the maximum distance in the experiment parameter space from any previously performed experiment. The value of S can be negative as the two distributions being used within the KL-divergence are not guaranteed to be equal.

The reasons for this trade-off are that when an observation is obtained that is not surprising, so agrees with the current hypotheses, we can infer that the confident hypotheses under consideration agree and a good representation of the behaviour for the features discovered has been made. If a good representation exists for the data available, then features of the behaviour not yet discovered should now be sought after through exploration. Whilst when an observation is surprising, the hypothesis manager will ensure that there are hypotheses that will have opposing views of the surprising observation, meaning that an exploitation experiment can be performed to identify the hypotheses that make the correct assumption about the surprising observation. It may be that several exploitation experiments are performed in a row that investigate one particular feature repeatedly, to allow for refinements of the hypotheses to be made. Once the most confident hypotheses provide a representation that describes that feature well, the observations will become unsurprising again and exploration will occur. Another way to consider this surprise technique is that exploitation will occur whilst the information gain is increasing to continue to obtain the information available, where information gain is measured through monitoring the change in prior and posterior confidences across the KL-divergence. When the information gain stops increasing, the technique will explore to try and obtain new sources of information.

The process of experiment selection occurs as follows. On the final experiment of the initial experiments and all subsequent experiments, the surprise of the observation obtained is calculated. After the surprise of the observation has been calculated, the hypothesis manager updates and refines the hypotheses using the process described previously. If S is positive, meaning the observation was surprising, then then next experiment will be an exploitation experiment chosen as the maximum of $D(x)$. If S is not positive, then the observation was not surprising and the next experiment will be an exploration experiment, chosen as the experiment that is maximally away from all other previously performed experiments in the experiment parameter space.

4.2. Selecting Experiments at Peaks of the Discrepancy Equation

As an alternative active strategy for the exploration–exploitation trade-off, we consider a strategy of combined exploration and exploitation based on the discrepancy between hy-

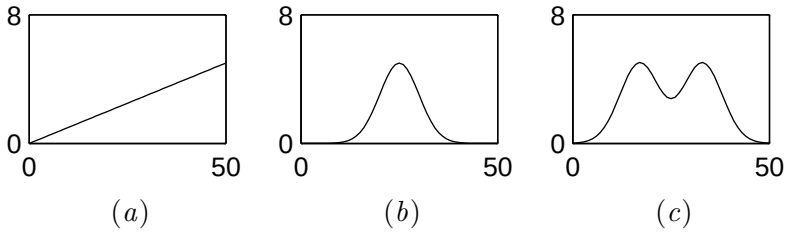

Figure 2: Underlying behaviours used to evaluate the artificial experimenter within a 1-dimensional experiment parameter space.

potheses. The exploitation function $D(x)$ in Equation 3, will give a maximal value where the hypotheses most disagree. Performing these experiments when there are good hypotheses in consideration, will identify the hypothesis that most suitably describes the underlying behaviour by disproving the alternate hypotheses. However these exploitation experiments will likely focus on particular areas of the parameter space and may place experiments close to each other in the parameter space. This will mean that little exploration will occur and features of the behaviour may be missed, or only a small number of the differences between hypotheses are examined.

Instead if we consider $D(x)$ across the parameter space, we may expect to see local maxima, or peaks, in the function. Each of these peaks should provide an area of the parameter space where the hypotheses disagree, potentially for different features of the behaviour. If experiments are placed at these peaks, there are three potential benefits. First, there will be a guaranteed information gain through identifying a difference between the hypotheses. Second, different differences between hypotheses will be examined. Finally, experiments will be placed across the parameter space allowing for some additional exploration.

The process of this technique is as follows. After building the initial set of hypotheses, a set of experiments are then chosen that are at the peaks of the discrepancy equation $D(x)$. These experiments are performed sequentially, with the hypotheses updated after each experiment is performed. When all experiments in the set have been performed, the discrepancy equation is recalculated and the next set of experiments are selected and performed.

5. Evaluation in 1-dimensional Parameter Space

To perform a simulated evaluation of the experiment selection techniques, we consider that characterisation experimentation can be described as a function:

$$y = f(x) + \epsilon + \phi \tag{11}$$

where y is the observation, x is the parameter for the experiment to perform on some system described by f, ϵ is observational noise present in all experiments and ϕ is shock noise present only in experiments that yield erroneous observations.

In previous work we demonstrated that selecting experiments at the peaks of the discrepancy equation (referred from here as discrepancy peaks) was consistently one of the best

strategies to perform in a 1-dimensional problem (Lovell et al., 2010). However in the more complex nonmonotonic behaviours, random selection of observations performed similarly to the discrepancy peaks technique. Using the same methodology described in that work, we evaluated the artificial experimenter on a set of simulated behaviours in a 1-dimensional problem (further details of this trial in 1-dimension can be found in Lovell et al. (2011)). To do this, 15 experiments were performed per trial, where the first 5 experiments were equally spaced across the experiment parameter space and the remaining 10 were actively selected using the technique being examined. Five initial experiments are chosen to allow a reasonably diverse set of initial hypotheses to be created. Of those 15 experiments, all observations were adjusted with Gaussian noise $\epsilon = N(0, 0.5^2)$ and 3 experiments produced erroneous observations by applying additional noise from $\phi = N(3, 1)$. The mean squared error between the most confident hypothesis and the true underlying behaviour for the set of possible experiment parameters were then calculated after each experiment, across 100 trials of each strategy using:

$$\begin{aligned} E &= k \int_{\mathcal{X}} \Big(\hat{b}(x) - f(x)\Big)^2 \, dx, \\ &= \frac{1}{|\mathcal{X}|} \sum_{n=1}^{|\mathcal{X}|} \Big(\hat{b}(x_n) - f(x_n)\Big)^2 \end{aligned} \tag{12}$$

where $\hat{b}(x)$ is the prediction of the most confident hypothesis for the experiment parameter x, which are drawn from the set of possible experiment parameters $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$. Importantly the set of hypotheses that b is chosen from is not complete and will change over time such that the set may not contain a good representation of the underlying function on a particular evaluation, making the problem not one of simply selecting experiments where the hypotheses maximally disagree, which would be appropriate if a good hypothesis was guaranteed within the set of possible hypotheses.

The behaviours used, shown in Fig. 2, tested the ability of the artificial experimenter to build models of behaviours representative of possible biological responses. The results, as shown in Fig. 3, demonstrate that the surprise technique was able to outperform the other techniques across the behaviours tested. Additionally, by performing a two-tailed t-test with 95% confidence tabulated in Table 1, the surprise technique is shown to provide statistically significant improvements over random selection in all cases and over the discrepancy peaks technique in most.

5.1. Laboratory Evaluation of Surprise Experiment Selection

Further to the simulated trial, a laboratory characterisation of the coenzyme NADH was performed. The coenzyme NADH was chosen for the trial as it is often used to monitor enzyme catalytic activity and the response can in-part be compared to the theoretical Beer-Lambert law in the linear region of the response (Nelson and Cox, 2008). The results for this evaluation comparing the surprise and discrepancy peaks experiment selection techniques are shown in Fig. 4. In each trial 5 initial exploratory experiments were performed followed by a further 10 actively selected experiments. A tabulation of whether the active experiments selected were exploration or exploitation experiments for the surprise technique are given in Table 2.

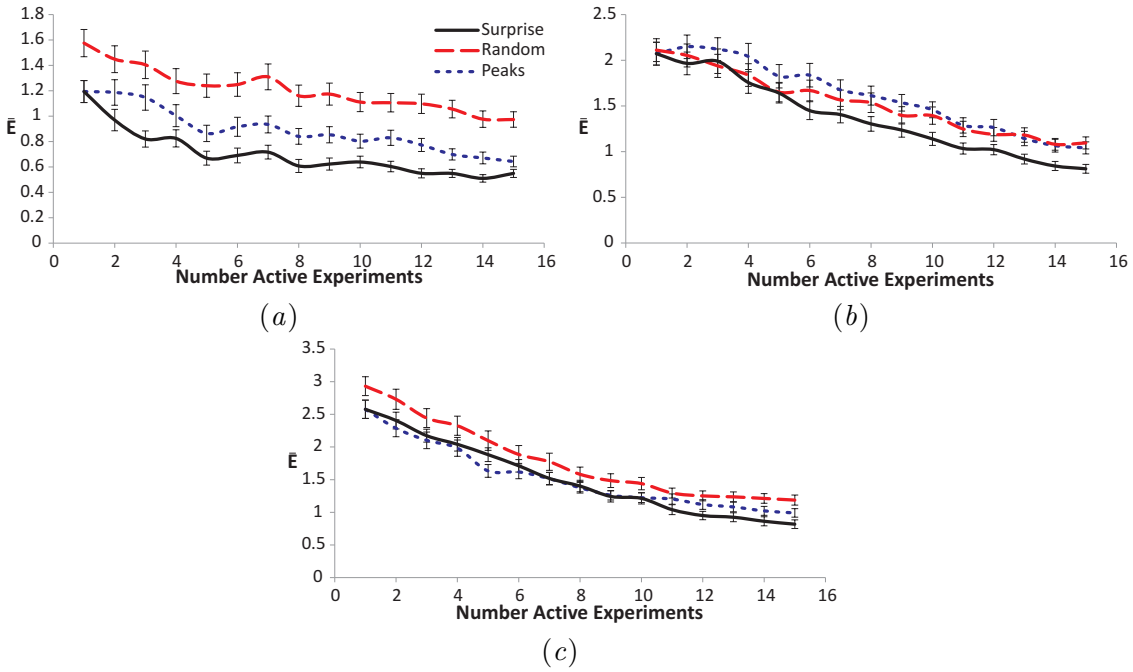

Figure 3: Mean error from 100 trials over number of active experiments performed, for the three behaviours and three experiment selection techniques being evaluated in the 1-dimensional parameter space evaluation. The error over time is shown for the active experiments that occur after the 5 initial experiments. In (a) the result for the behaviour shown in Fig. 2(*a*) is shown, (b) corresponds to Fig. 2(*b*) and (c) corresponds to Fig. 2(*c*). In each case the surprise technique outperforms the alternative techniques.

Behaviour	Technique	Active experiments with significant difference
A	Random	all
	Discrepancy	3, 5–14
	Peaks	
B	Random	10, 11, 13–15
	Discrepancy	6, 8–15
	Peaks	
C	Random	11–15
	Discrepancy	none
	Peaks	

Table 1: Identification of statistically significant results in the 1-dimensional evaluation. In each a comparison is made between the surprise technique and the one stated. In all cases the surprise technique provides a significant improvement over the alternate technique. The behaviours correspond to the ones shown in Fig. 2.

5.1.1. Surprise

After performing the initial exploratory experiments the artificial experimenter requested an experiment to be performed with concentration 0.42 mM, to examine the change in behaviour from a linear response in the lower concentrations. The observation obtained for that experiment agreed with the other experiment at 0.38 mM, causing the most confident hypotheses in that region to have a similar response. Next the experiment selection technique chose experiments to evaluate the region between 0.75 mM and 1.13 mM. In this region it found noisy observations, causing several exploitation experiments to be performed to investigate the different hypotheses within in this region. After 6 actively chosen experiments, the most confident hypothesis produced a good representation of response curve, with the initial linear component of the response prediction between 0 mM and 0.4 mM being similar to the theoretical prediction. After 8 actively chosen experiments, the response curve of the most confident hypothesis was essentially the same as it was after 6 actively chosen experiments, suggesting that too many exploitation experiments were performed at this stage of experimentation.

On the penultimate experiment the artificial experimenter performed an exploration experiment, as the extensive examination of the region between 0.75 mM and 1.13 mM ended with hypotheses that represented that region well, which caused the final experiment performed in that region to not be surprising. This exploration experiment at 0.18 mM produced an observation much higher than the hypotheses predicted, which caused the observation to be surprising. This surprising observation caused the final experiment to be an exploitation experiment to examine why the observation differed from the predictions of the hypotheses. The final experiment obtained an observation that agreed with the previous observation, causing the most confident hypotheses to believe that the behaviour passes through those observations and away from the Beer-Lambert law's theoretical prediction. This difference between the prediction and theoretical value should not be classed as a problem caused by the surprise technique, but rather due to the experimental observations obtained. It is likely that these two final observations were not representative of the true underlying behaviour, but because they both agreed with each other, the hypothesis manager believed the observations to be true.

5.1.2. Discrepancy Peaks

The discrepancy peaks technique initially chose experiments to examine near the stationary points of the curve, where the shape of the behaviour changes. In the region 1 mM and 1.3 mM, the observations obtained were fairly noisy, which resulted in continual exploitation of the the differences in hypotheses in this region. Throughout experimentation the technique also placed a large number of observations near a concentration of 0 mM. The repeated placement of experiments in this region is a weakness, as the observations were of similar absorbance measurements, with no new information being gained from the repeated trials. However, by the final active experiment, the technique had produced a good fit of the data, with a prediction of the linear region that was near identical to the Beer-Lambert theoretical prediction.

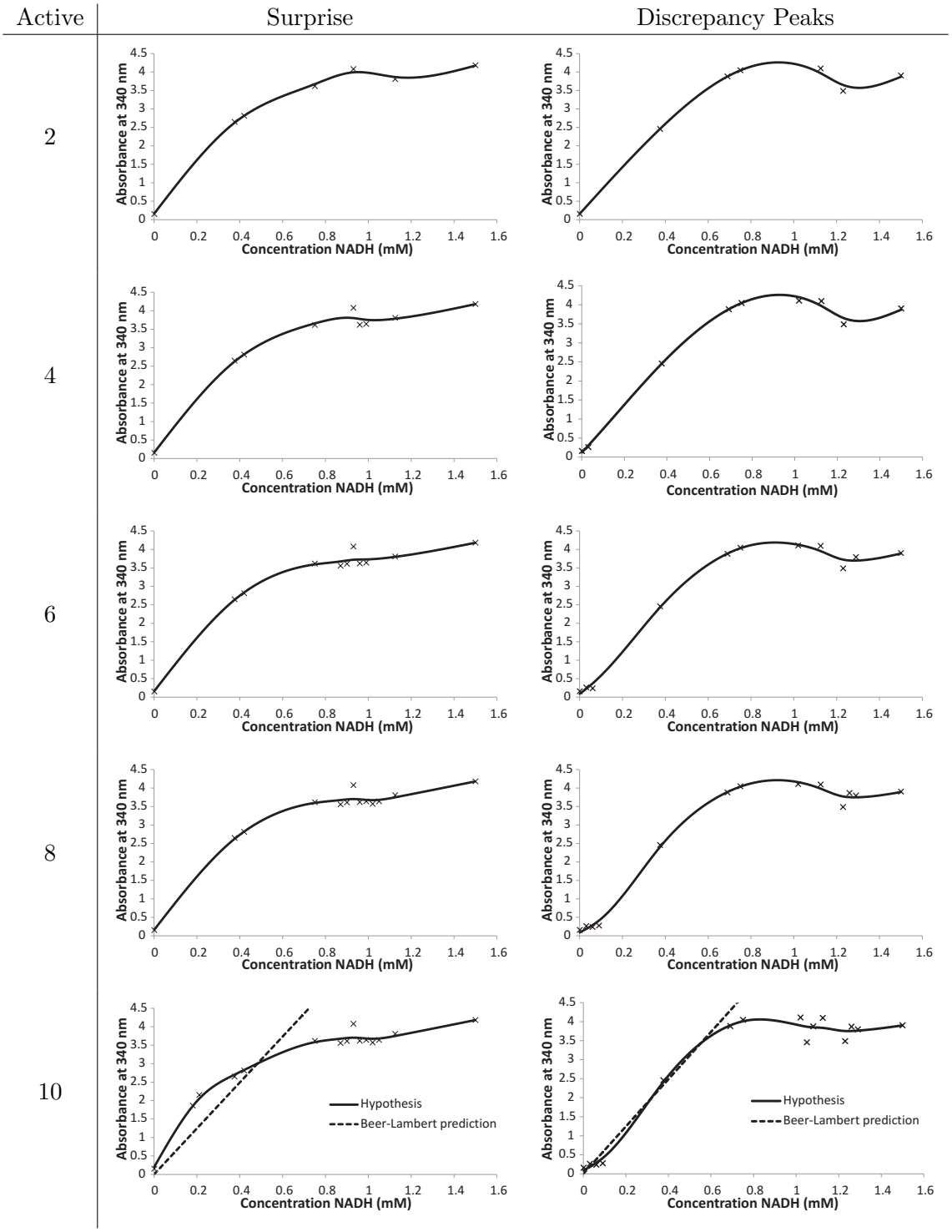

Figure 4: Most confident hypothesis over 10 actively chosen experiments for the discrepancy peaks and surprise explore-exploit switching experiment selection technique in a laboratory trial. Figures show intervals of 2 actively chosen experiments.

Active Experiment No.	Explore or Exploit
1	Exploit
2	Explore
3	Exploit
4	Exploit
5	Exploit
6	Exploit
7	Exploit
8	Exploit
9	Explore
10	Exploit

Table 2: Listing of whether surprise explore-exploit switching technique chose an exploration or exploitation experiment to perform.

5.1.3. Comparison

Both techniques first used experiments to investigate the region of the parameter space where the behaviour curves, roughly between 0.6 mM and 1.3 mM. In both trials, noisy observations were observed in the region 0.8 mM to 1.3 mM, leading to both techniques focussing experiments in this region. The surprise technique was able to determine quicker than the discrepancy peaks technique that this noisy region had an underlying behaviour that was roughly linear. This was due to the discrepancy peaks technique placing repeated experiments near 0 mM, caused by exploiting small differences between hypotheses.

Whilst the discrepancy peaks technique failed to allow much exploration of the parameter space in this trial, the experiment parameters it chose for the majority of experiments were in good regions of the parameter space to identify the features of the behaviour. In contrast the surprise technique identified the behaviour quicker and was able to perform further exploration. In the surprise trial it was only due to the final two observations being erroneous and in agreement, which led to the final hypothesis to be produced that incorrectly characterised the part of the behaviour that could be theoretically defined. Overall it appears that the surprise technique provided the better strategy in this trial, because it was able provide data that led to agreement about the noisy region quickly, allowing it to redirect resources back to exploration to search for features of the behaviour not yet characterised.

5.1.4. Materials and Methods

A stock solution of 5 mM NADH and a 10 mM Tris buffer at pH 8.5 were prepared. Dilutions of NADH requested by the artificial experimenter were produced by mixing volumes taken from the stock solution and the buffer. Measurements of optical absorbance at 340 nm were recorded with a PerkinElmer Lambda 650 UV-Vis Spectrophotometer to provide the observations. The photometric range of the spectrophotometer was 6 A.

6. Evaluation in 2-dimensional Parameter Space

We now evaluate the experiment selection techniques within a 2-dimensional parameter space. In each case the multiple hypotheses based hypothesis manager was used, combined with either the random, discrepancy peaks or surprise based technique for experiment selection. The protocol for the hypothesis manager remained the same as the 1-dimensional version, except that for performance 40 new hypotheses were created in each iteration (down from 200 in the 1-dimensional version), with the best 100 hypotheses kept into the next round of experimentation (where the top 20% of the hypotheses were kept on each iteration in the 1-dimensional version). Hypotheses were represented using a thin plate spline (Wahba, 1990):

$$h = \min \sum_{i,j}^{n} (y - f(x_1, x_2))^2 + \lambda \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(f''(x_1, x_1)^2 + 2f''(x_1, x_2)^2 + f''(x_2, x_2)^2 \right) dx_1 dx_2 \tag{13}$$

with a choice of smoothing parameters ($\lambda \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$). All independent parameters were coded between 0 and 1, from behaviours with uncoded x_1 and x_2 parameters both ranging between 0 and 50. The underlying behaviours used are presented in Fig. 5, where (a) provides a single feature, (b) a behaviour where only the x_2 factor provides a role in determining the response, and (c) a behaviour with two peaks and a trough. In each case the behaviours were between 0 and 8 on the dependent variable, so that the noise parameters ϵ and ϕ could remain the same for both the 1 and 2-dimensional evaluations.

The three experiment selection strategies were tested over 100 trials per behaviour. In each behaviour, 5 initial experiments were performed, which were equally spaced around the parameter space ([0,0], [1,0], [0,1], [1,1] and [0.5,0.5] in coded values). After the exploratory experiments, a further 25 actively chosen experiments were performed per trial, where 3 of the experiments produced erroneous observations. Gaussian noise was added to all observations with $\epsilon = N(0, 0.5^2)$, whilst the noise applied to erroneous observations was $\phi = N(3, 1)$. The techniques were again evaluated by comparing the mean over 100 trials of the error between the most confident hypothesis and the true underlying behaviour.

6.1. Results

In the 2-dimensional problem, the results show there is less difference between the surprise and random experiment selection techniques than in the 1-dimensional case, whilst the discrepancy peaks technique again generally performs the worst, as shown in Fig. 9. However, overall it appears that the surprise technique is still a more robust technique than the others considered, with the technique providing significant improvements over a random strategy in two of the three underlying behaviours. In Fig. 6, 7 and 8, a comparison of most confident hypotheses for each technique and behaviour after 25 active experiments are shown. In each case the error between the hypothesis shown and the true underlying behaviour is representative of the mean error given in Fig. 9.

For the single feature behaviour, A, the random technique outperforms the surprise technique between the 7th and 23rd active experiments, as shown in Fig. 9(*a*). During

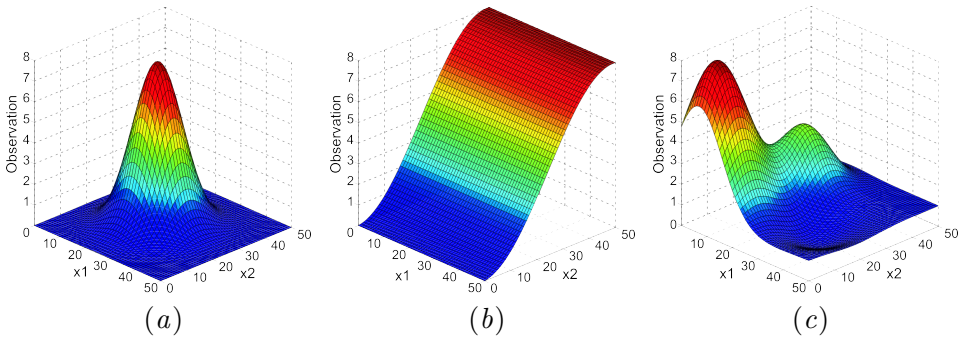

Figure 5: Underlying behaviours used for simulated 2-dimensional parameter space trials.

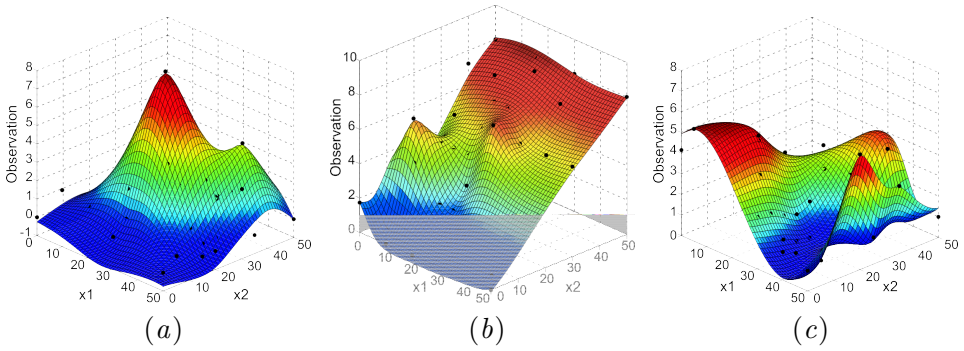

Figure 6: Representative hypotheses produced by random experiment selection.

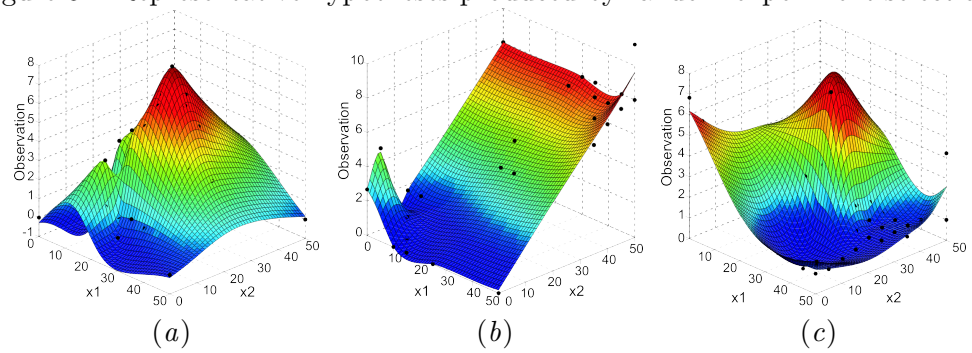

Figure 7: Representative hypotheses produced by discrepancy peaks experiment selection.

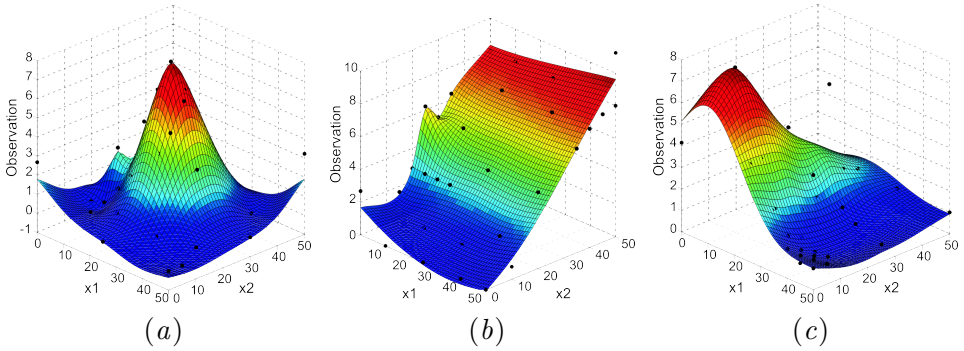

Figure 8: Representative hypotheses produced by surprise experiment selection.

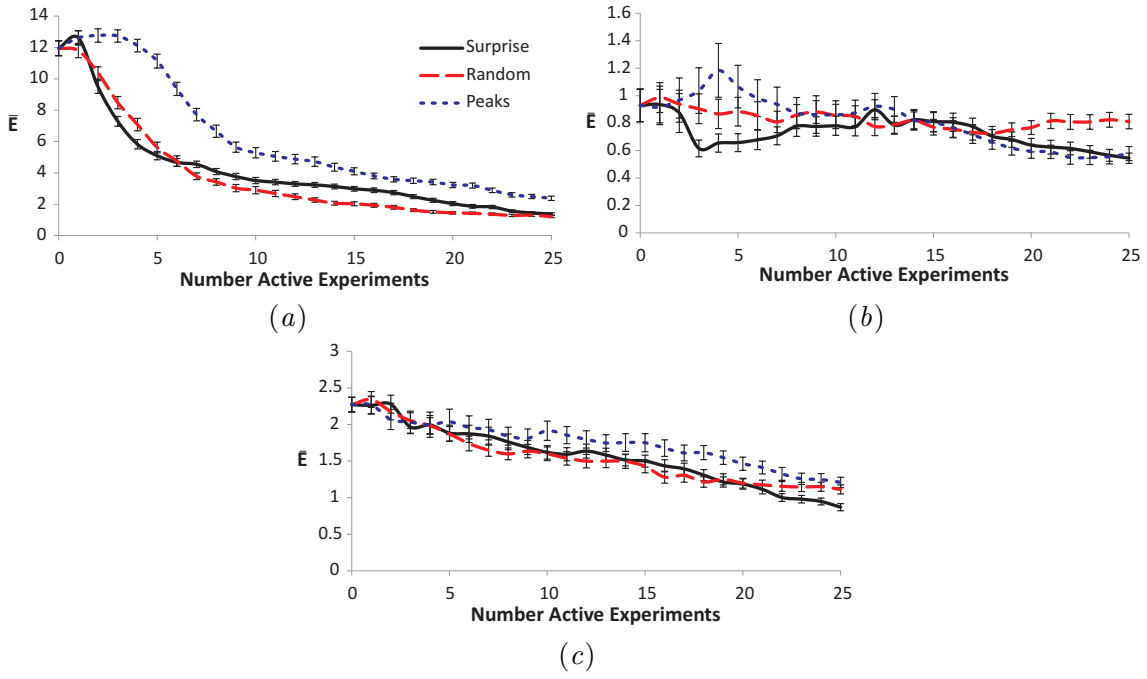

Figure 9: Mean error from 100 trials over number of active experiments performed, for the three behaviours and three experiment selection techniques being evaluated in the 2-dimensional parameter space evaluation. The error over time is shown for the active experiments that occur after the 5 initial experiments. In (a) the result for the behaviour shown in Fig. 5(a) is shown, (b) corresponds to Fig. 5(b) and (c) corresponds to Fig. 5(c).

these experiments, the surprise technique spends more time investigating smaller differences between the hypotheses, causing a greater amount of exploitation early on than is perhaps necessary. Whilst the random technique is able to explore the parameter space more early on, allowing it to form a better general understanding of the behaviour quicker than the surprise technique. However, the random technique can perform poorly if it samples a region of the parameter space only once and obtains an erroneous observation, which can cause it to include an additional feature in the prediction that is not present in the underlying behaviour, as shown in Fig. 6(a). The discrepancy peaks technique performs poorly throughout, by over exploiting the information obtained rather than exploring. This causes discrepancy peaks technique to continually investigate small differences between the hypotheses, caused by the Gaussian noise applied to each observation.

For the single factor behaviour, B, the initial 5 data points, if error free, are capable of providing all of the techniques with data suitable for producing a good representation of the behaviour. Therefore this behaviour tests the ability of the experiment selection techniques to deal with erroneous observations in a 2-dimensional parameter space. The random technique fails to improve the performance of the most confident hypothesis throughout the 25 actively chosen experiments. In part this is caused by the technique not investigating

erroneous observations, so any improvements to understanding the behaviour are lost by erroneous observations that misguide hypothesis formation. The surprise technique performs well early on, as it is able to investigate erroneous observations. However the mean error increases again between 7 and 15 experiments because the technique over investigates some observations, causing the hypotheses to overfit some of the noise. The discrepancy peaks technique also suffers the problem of over sampling a region, causing lots of hypotheses with differences of opinion in a small area, which leads to hypotheses overfitting in those areas the ϵ noise applied to the observations. However, over time the discrepancy peaks technique lowers the error to slightly below the surprise error in the latter stages of experimentation. By 25 experiments the performance of the surprise and discrepancy peaks techniques are nearly equal.

For the behaviour with multiple features, C, the resources available were too few to get a good representation of the behaviour. The surprise and random techniques reduce the error at a similar rate for the first 18 active experiments. However after 18 active experiments the error for the random technique levels out, whilst it continues to reduce for the surprise technique. In the random technique the experiments are spread out across the parameter space, allowing for the different features to be identified quickly, albeit at a low resolution. However, as the experiments are not directed, increasing the understanding of any particular behaviour is by chance and potentially erroneous observations are ignored. These two factors prevent the error from reducing further later on in experimentation. The surprise technique through performing exploitation experiments, performs more experiments near the features it discovers, causing better representations of the behaviour to be formed. Additionally, the technique is able to investigate and identify erroneous observations, whilst also performing experiments to further search the parameter space. The discrepancy peaks technique performs worse than the other two techniques, because it over exploits and becomes focussed in particular regions of the parameter space where the first unexpected behaviours were obtained. Unlike the other two behaviours where all techniques provided a somewhat representative prediction of the underlying behaviour, the surprise based technique is the only technique to provide a good representation of underlying behaviour C, shown in Fig. 5(c), 6(c), 7(c) and 8(c).

Like the 1-dimensional case, the results for the 2-dimensional parameter space have been analysed using a two-tailed t-test with $\alpha = 0.05$ to determine if the results are significant at the 95% confidence interval, shown in Table 3. The surprise technique provided significant improvements over a random selection strategy for behaviours B and C, although only in the latter stages of experimentation. These improvements are in part due to the surprise technique being able to better identify erroneous observations than the random technique. Additionally the surprise technique is able to investigate the new features it discovers further, which allows it to provide a better representation of the more complex behaviour C. The random strategy performs significantly better than the surprise technique for the majority of the experimentation performed using behaviour A. This is due to the random technique being able to explore the parameter space more, where the surprise technique spends some additional time investigating small differences between the hypotheses that only provide small benefits for developing a representation of the behaviour. The discrepancy peaks technique performs significantly worse than the surprise technique for behaviours A and C by the end of the experimentation performed. However, the technique

Behaviour	Technique	Active experiments with significant difference
A	Random	3, 4, (7–23 surprise is significantly worse)
	Discrepancy Peaks	2–25
B	Random	3, 21–25
	Discrepancy Peaks	3–6
C	Random	23–25
	Discrepancy Peaks	18–25

Table 3: Identification of statistically significant results in the 2-dimensional evaluation. In each a comparison is made between the surprise technique and the one stated. Except where stated, the surprise technique provides a significant improvement over the alternate technique. The behaviours correspond to the ones shown in Fig. 5.

performs similarly well compared to the surprise technique for the behaviour B, due to little exploration being required, which the technique is poor at providing.

7. Related Work

The exploration vs. exploitation trade-off, although not always specifically mentioned, has been addressed by many of the computational scientific discovery systems in the literature. In this section we first provide a brief review of the techniques employed in these systems to address the exploration–exploitation trade-off. Then we consider active learning techniques and their current limitations within autonomous experimentation.

The KEKADA system was one of the first examples of an artificial experimenter, which formed more structural hypotheses about the mechanisms of reactions (Kulkarni and Simon, 1990). The system attempted to follow the heuristics used by Hans Krebs to discover the urea cycle, to determine whether a computational system could rediscover knowledge obtained in physical experimentation. A key part of the KEKADA system was reacting to surprising observations, which are those observations that do not agree with the current view of the behaviour being investigated. When a surprising observation was obtained, the system would follow several strategies for reacting to the surprise that would exploit the information available. Examples of the strategies employed are: identifying the independent parameter that caused the surprise and identifying errors in the current hypotheses. The system was able to provide a good model of the heuristics used by Hans Krebs, however it could be easily outperformed by human experimenters who had far more heuristics available to them.

The FAHRENHEIT system, an extension of BACON (Langley et al., 1987), was designed to find an empirical theory that could describe an observed behaviour within a parameter space (Żytkow et al., 1990). FAHRENHEIT was demonstrated to work in an autonomous experimentation machine in the field of electro-chemistry (Żytkow et al., 1990). To begin the system performed experiments that explored the parameter space, to allow it to produce models of regularities in the behaviour. When the system discovered an irregularity, for example a rapid change in phase such as a discontinuity, the system would then focus

experiments on investigating the extent of the irregularity, producing separate models for the regions of regularity adjacent to the irregularity. However this system would require a large number of experiments to be performed to provide the data required.

A system developed to automate a chemistry workstation, employed a grid search with decreasing grid size as part of its strategy to manage the exploration–exploitation trade-off (Dixon et al., 2002). The goal of the system was to discover the parameters that produced the highest yield within the experiment parameter space. Initially experiments were placed spread out across the parameter space to provide exploration by a grid with large grid squares. The size of the grid squares decreased over subsequent experiments, to provide a more detailed analysis of the behaviour. Additionally a simplex based experiment selection technique was employed in later stages of experimentation, where experiments would focus towards areas of the parameter space where previously high yields were obtained (Du and Lindsey, 2002). Like many evolutionary algorithms, the technique had the potential for becoming stuck within a local maxima.

Scouting was an evolutionary algorithm that evolved parameters based on an adaptive measure of surprise, which was able to manage the trade-off between exploration and exploitation (Pfaffmann and Zauner, 2001). Like KEKADA, surprising observations were those that differed from the hypothesis under consideration. When no observations were surprising, experiments would be placed randomly within the parameter space. When a surprising observation was obtained, the evolutionary algorithm would then place experiments near the surprising observation. Importantly, as more experiments were placed near the initially surprising observation, so the model would better represent the behaviour and the observation would become less surprising. This adaptivity of surprise, meant that once sufficient information had been obtained to investigate why the observation was surprising, the algorithm would again place experiments in other areas of the parameter space, automatically addressing the exploration-exploitation trade-off. The scouting approach was demonstrated within an autonomous experimentation machine to characterise enzymatic response behaviours (Matsumaru et al., 2002). A problem with this technique was that if an observation was erroneous, then it could remain surprising as subsequent observations would not agree with it, meaning that the system could remain performing exploitation experiments in that region.

The robot scientist (King et al., 2004) does not algorithmically consider the trade-off between exploration and exploitation. Instead the system is provided with a large body of information within a relatively small domain, which is then used to formulate hypotheses that can be tested to determine their validity. Essentially the initial information provided by the user is the only exploration that occurs. The active learning technique the system uses to select experiments to examine the hypotheses is purely exploitative, by choosing experiments that will reduce the likely cost to determine the most representative hypothesis.

Another technique that investigates automatically characterising enzymatic response characterisation, performs a largely exploration focussed experiment selection (Bonowski et al., 2010). Initially the technique is explorative through placing experiments using a space fitting algorithm to ensure a good distribution of experiments across the experiment parameter space. Later experiments combine exploitation and exploration through placing experiments where the uncertainty in the hypothesis is greatest, but also explorative through requiring experiments to fulfil a minimum distance requirement between experiments in

the parameter space. The minimum distance required decreases over time, to allow finer examination of the parameter space. However the technique does not consider erroneous observations.

7.1. Active Learning

Active learning seeks to address the same issues as those in autonomous experimentation. That is to minimise the number of labels to obtain, or experiments to perform, whilst maximising the information obtained. In comparison to present autonomous experimentation techniques, active learning is more theoretically grounded and mathematically sound than the often ad-hoc techniques found in autonomous experimentation. However at present the theoretical assumptions made in active learning mean the problems addressed are not always representative of physical experimentation, which limits their potential applications. In particular the assumption that experiments occur without noise, is an assumption that is made too frequently (Cohn et al., 1994; Freund et al., 1997), or that a hypothesis is already in consideration that provides a suitable representation of the behaviour being investigated (MacKay, 1992; Settles, 2009).

The work by MacKay (1992) places experiments where the predicted information gain is highest, either in a single hypothesis or to discriminate between multiple models. In order to do this, the assumption is made that at any particular point in the experimentation, a model exists that is representative of the underlying behaviour being investigated. This makes experiment selection purely exploitation based and is honing in on the most appropriate model or hypothesis. In physical experimentation this will often not be the case. Whilst the method for discriminating between hypotheses proposed by MacKay is useful in autonomous experimentation, and is similar to the discrepancy function we use in Eqn. 3, the technique as a whole cannot be used as is within physical experimentation, due to the assumption about a complete model space.

Query-by-committee provides a way of managing the uncertainty in predictions through allowing an ensemble of hypotheses to be considered in parallel, similar to how the multiple hypotheses are used within the approach considered here (Seung et al., 1992). The technique takes an ensemble of hypotheses and performs experiments in the locations where there is the maximal disagreement between hypotheses. The disagreement is considered maximal where there are equal votes for the two options within a binary classification problem. However the query-by-committee is stated to have limitations that restrict the use of the algorithm within practical applications, with the primary limitation being the assumption that experiments are noise free (Freund et al., 1997). The idea of considering multiple hypotheses is important in managing the uncertainty presented within autonomous experimentation problems, therefore query-by-committee is a highly appropriate technique to apply within such active learning tasks. Although alterations are required to the technique in terms of hypothesis management and discrepancy calculation, like those presented here, before this technique can be more widely used within autonomous experimentation.

Active learning has considered the possibility of erroneous observations, described in the literature as noisy oracles (Settles, 2009). However research into this particular area is currently extremely limited and do not address the issue found in experimentation where erroneous observations occur sporadically, with no particular distribution or consistency.

8. Conclusion

In physical experimentation, the costs per experiment prevent large numbers of observations from being obtained. With the validity of observations not guaranteed, experiments must be performed that test the validity of the hypotheses produced and also search the parameter space under investigation for features of the behaviour not yet discovered. This trade-off between feature discovery and hypothesis evaluation is an exploration–exploitation trade-off, where differences between competing hypotheses is used in the exploitation. The resources available prevent a large number of repeated experiments in an area to get a highly accurate prediction. Likewise the resources limit the exploration that can be performed to find all the different features of the behaviour. Techniques are therefore required that will search for features of the behaviour under investigation, whilst ensuring a reasonable level of confidence that the hypothesised feature observed is genuine.

To manage this trade-off we consider a technique that is similar to how successful human experimenters address the problem in physical experimentation, by using the surprise of the last observation obtained to determine whether the next experiment will explore or exploit. A Bayesian formulation of surprise has been used, where an exploration experiment is performed when the last experiment was not surprising and an exploitation experiment is performed when the last experiment was surprising. This use of surprise ensures that experiments are performed to evaluate hypotheses when an observation is discovered that the hypotheses did not expect, to determine why the hypotheses did not expect that observation, either due to the observation being erroneous or the hypotheses inaccurate. Whilst when observations are obtained that are similar to the predictions of the most confident hypotheses under consideration, then exploration is performed to look for features of the behaviour not yet captured by the hypotheses.

Expanding to higher dimensions, the surprise technique was in some cases able to provide a significant improvement over the alternate techniques considered. However the degree of benefit was less than in the 1-dimensional case. A limitation of the surprise technique in the higher dimension problems, was that it performs little exploration early on. As the thin plate splines tended to overfit observation noise more than the smoothing splines did with 1-dimensional data, there was greater discrepancy between the hypotheses, which caused the surprise technique to exploit more often, especially in early experiments where most observations would be surprising to the hypotheses. This led to some of the early experiments being focussed within a particular area. However, unlike the discrepancy peaks technique that could also focus the placement of experiments within a small area, the surprise technique adapted over time and explored the parameter space, leading to it providing more accurate hypotheses than the other techniques later on in experimentation. The surprise technique could benefit from additional initial exploration, however care would have to be taken to ensure that this exploration does not bias particular regions of the parameter space.

Acknowledgments

The reported work was supported in part by a Microsoft Research Faculty Fellowship to KPZ. This work was supported in part by the IST Programme of the European Community, under the PASCAL2 Network of Excellence, IST-2007-216886. This publication only reflects the authors' views.

References

P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3:397–422, 2002.

F. Bonowski, A. Kitanovic, P. Ruoff, J. Holzwarth, I. Kitanovic, V. N. Bui, E. Lederer, and S. Wölfl. Computer controlled automated assay for comprehensive studies of enzyme kinetic parameters. *PLos ONE*, 5(5):1–10, 2010.

T. C. Chamberlin. The method of multiple working hypotheses. *Science (old series)*, 15: 92–96, 1890. Reprinted in: Science, v. 148, p. 754–759, May 1965.

D. Cohn, L. Atlas, and R. Ladner. Improving generalization with active learning. *Machine Learning*, 15:201–221, 1994.

J. M. Dixon, H. Du, D. G. Cork, and J. S. Lindsey. An experimental planner for performing successive focused grid searches with an automated chemistry workstation. *Chemometrics and Intelligent Laboratory Systems*, 62:115–128, 2002.

H. Du and J. S. Lindsey. An approach for parallel and adaptive screening of discrete compounds followed by reaction optimization using an automated chemistry workstation. *Chemometrics and Intelligent Laboratory Systems*, 62:159–170, 2002.

Y. Freund, H. S. Seung, E. Shamir, and N. Tishby. Selective sampling using the query by committee algorithm. *Machine Learning*, 28:133–168, 1997.

L. Itti and P. Baldi. Bayesian surprise attracts human attention. *Vision Research*, 49: 1295–1306, 2009.

R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant, S. H. Muggleton, Douglas B. Kell, and Stephen G. Oliver. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*, 427:247–252, 2004.

D. Kulkarni and H. A. Simon. Experimentation in machine discovery. In J. Shrager and P. Langley, editors, *Computational Models of Scientific Discovery and Theory Formation*, pages 255–273. Morgan Kaufmann Publishers, San Mateo, CA, 1990.

P. Langley, H. A. Simon, G. L. Bradshaw, and J. M. Żytkow. *Scientific Discovery: computational explorations of the creative processes*. MIT Press, 1987.

C. J. Lovell, G. Jones, S. R. Gunn, and K.-P. Zauner. An artificial experimenter for enzymatic response characterisation. In *13th International Conference on Discovery Science*, pages 42–56, Canberra, Australia, 2010.

C. J. Lovell, G. Jones, S. R. Gunn, and K.-P. Zauner. Autonomous experimentation: Active learning for enzyme response characterisation. *JMLR: Workshop and Conference Proceedings*, 16:141–154, 2011.

D. J. C. MacKay. Information–based objective functions for active data selection. *Neural Computation*, 4:589–603, 1992.

N. Matsumaru, S. Colombano, and K.-P. Zauner. Scouting enzyme behavior. In D. B. Fogel, M. A. El-Sharkawi, X. Yao, G. Greenwood, H. Iba, P. Marrow, and M. Shackleton, editors, *2002 World Congress on Computational Intelligence, May 12-17*, pages CEC 19–24, Honolulu, Hawaii, 2002. IEEE, Piscataway, NJ.

D. L. Nelson and M. M. Cox. *Lehninger Principles of Biochemistry*. W. H. Freeman and Company, New York, USA, 5th edition, 2008.

J. O. Pfaffmann and K.-P. Zauner. Scouting context-sensitive components. In D. Keymeulen, A. Stoica, J. Lohn, and R. S. Zebulum, editors, *The Third NASA/DoD Workshop on Evolvable Hardware—EH-2001*, pages 14–20, Long Beach, July 2001. IEEE Computer Society, Los Alamitos.

B. Settles. Active learning literature survey. Technical report, University of Wisconsin-Madison, 2009.

H. S. Seung, M. Opper, and H. Sompolinsky. Query by committee. In *Proceedings of the ACM Workshop on Computational Learning Theory*, pages 287–294, 1992.

G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference series in applied mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.

J. M. Żytkow, J. Zhu, and A. Hussam. Automated discovery in a chemistry laboratory. In *Proceedings of the 8th National Conference on Artificial Intelligence*, pages 889–894, Boston, MA, 1990. AAAI Press / MIT Press.