Getting a monkey to do your bidding

Developing a Becker-DeGroot-Marschak (BDM) method for use in monkeys

Alaa Al-Mohammad

Trinity Hall

September 2017

This dissertation is submitted for the degree of Doctor of Philosophy

Contents

Preface

This dissertation was completed in the department of Physiology, Development and Neuroscience under supervision of Professor Wolfram Schultz.

The following dissertation is a result of my own work and includes nothing which is the outcome of work done in collaboration, except where specified in the text. It is not substantially the same as any other work that I have submitted, or, that is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other university or similar institution.

Furthermore, no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge, or, at any other institution.

This dissertation does not exceed the prescribed word limit for the Biology Degree Committee.

Acknowledgements

The work presented here was made by possible by the help of friends and colleagues. First, I want to thank Professor Wolfram Schultz for his constant support and trust throughout the four years of research.

I would also like to thank all members of Professor Schultz's lab, particularly: Dr Armin Lak, Dr Bill Stauffer, Dr Raymundo Baez-Mendoza, and Dr Alexandre Pastor-Bernier, for their early enthusiasm for my project and suggestions for improvements and possible future directions; Dr Charlotte van Coeverden, Dr Kelly Diederen, Dr Martin Vestergaard, Dr Leopold Zangemeister and Philipe Bujold for interesting theoretical discussions and moral support when the monkeys were misbehaving; Dr Fabian Grabenhorst for always being there to answer my questions and suggest interesting analyses; Aled David and Christina Thompson for making sure that I could sleep well knowing that my monkeys were in good hands.

My friends were also a constant source of support. Specifically, I would like to thank Dr Louis Nerurkar, who was with me every step of the way, and always knew what I was going through because he was going through it himself. Most importantly, I want to thank my family who have sacrificed so much and who were there for me always and unconditionally: my parents, Abdallah Al-Mohammad and Maysoon Kaddoura, and my siblings, Deema, Abdulrahman, and Leen.

Summary

The Becker-DeGroot-Marschak method (BDM) is an auction-like mechanism widely used in behavioural economics, marketing research, and, more recently, in neuroimaging studies of human decision making. The BDM has never been used with animal subjects before, yet its application in monkeys would allow for comparison of studies across species while providing a direct measure of what a reward is worth to a monkey in a single experimental trial.

In the BDM, a subject is given a budget with which they can place a bid for some reward, and a computer then randomly selects a competing bid. If the subject's bid is higher than the computer's bid then the subject pays an amount equal to the computer's bid, receives the reward object, and gets to keep the remaining budget. If the subject's bid is lower than the computer's bid, the subject does not gain the reward object but retains the entire budget.

To adapt the task for monkeys, two rhesus macaques were taught to use water as a budget, and to use a joystick to place a bid in terms of this budget for different volumes of fruit-juice reward.

The BDM ensures that the subject's optimal action is to place a bid equal to their value for the reward-object. This property of truthful value revelation is the BDM's most important feature in the context of value-based decision making.

Currently, the only method of eliciting a monkey's value for one reward in terms of another depends upon inference of the magnitudes at which the two rewards are chosen with equal probability. Using this 'binary-choice' method, many trials are needed to infer a single value: pairwise comparisons of many different magnitudes must be made and choices of each pair must be repeated so that the probability of choosing a reward can be estimated.

In contrast, the BDM provides a direct measure of the monkey's value for the reward as they explicitly state this value on each trial by selecting an equivalent bid. Therefore, the BDM more efficiently utilises the limited time in which a monkey's behaviour can be assessed in each experimental session, as animals lose the motivation to participate when they become sated.

The thesis summarised here describes the training and performance of two rhesus macaques on a novel version of the BDM, specifically designed for a subject that cannot be instructed on the optimal strategy. The technical steps and intermediate tasks that are needed to train a monkey to flexibly place bids by operating a joystick are also detailed, as well as the development of different versions of the task over three years of testing. The results of the final version of the BDM are then presented for both monkeys, showing rational bidding behaviour consistent with an understanding of the method’s contingencies. Theoretical concerns and limitations of the BDM in such a setting are also discussed and the thesis outlines how future experiments can make use of and adapt this version of the BDM for neuronal recording experiments.

Glossary

Short definitions of technical terms that are often used throughout this thesis are provided here. Where possible, these concepts and terms have also been explained in the main text as they are introduced. Words italicised in the main text of this dissertation appear in this glossary:

Subjective value – The value of a reward as determined by the subject, independent of any inherent property of the reward. It is used here specifically to refer to the value that is placed upon one reward in terms of another – to draw a distinction with the concept of utility.

Numeraire – A reference or standard by which values are computed. Here it is used to refer to the reward in terms of which the values of other rewards are expressed.

Utility – A measure of preferences and satisfaction of the subject in entirely subjective terms.

Utility function – A mathematical function which ranks rewards in terms of their utility for the subject. It is used here to refer to cardinal utility functions, wherein the magnitude of the utility is a measure of the strength of preference.

Prediction Error – The difference between the expected outcome and the realised outcome of some event. The firing of dopamine neurons of the midbrain is known to reflect behaviourally relevant prediction errors, or PEs.

Good – Here this term is used specifically for the juice-reward that is to be bid for in the BDM.

Budget – Here this term is used specifically to refer to the water reward that the monkey uses to place bids, and therefore ‘spends’ in the BDM task.

Incentive compatibility – A method is incentive compatible if the participants’ best strategy is to act in accordance with their true preferences.

Weakly dominant strategy – A strategy is said to be weakly dominant if the subject's payoffs are at least as high as they would be under any other strategy regardless of what other subjects do, and, earns a strictly higher payoff than other strategies for some of the other subjects' strategies.

Indifference point – The point at which a subject has no preference between two different rewards and will choose them in equal proportion given repeated choices between them.

Bandit Task – Or 'multi-armed bandit' is a problem in which different options produce random rewards with a specific underlying probability distribution.

Lottery – A reward with several different probabilistic outcomes. This is often called a 'risky' reward in the neuroscience of decision making literature.

Certainty Equivalence – The guaranteed payoff for which the subject is indifferent between it and a lottery, or risky reward.

Fractile method – A method whereby certainty equivalents are found for different lotteries in order to map the subject's utility function.

Induced value – An induced value bidding procedure involves offering subjects tokens with specific known monetary values for which they can be traded following completion of the experiment. Induced value experiments are used to assess the performance of human subjects in various bidding tasks.

Probability distortion – The phenomenon of non-linear subjective representation of probabilities by both human and animal subjects.

Endowment effect – The hypothesis that subjects place a higher value on objects that they already own.

First-price auction – An auction whereby the highest bidder wins and pays an amount equal to their bid.

Income and substitution effects – The effect of a change in prices on a subject's disposable income and their choice of alternative goods, or in turn, the effects of a subject's income and the availability of alternative goods on the value that they place on some other reward.

Temporal discounting – The observed tendency of subjects to discount the value of rewards the further into the future that they expect to receive them.

Bid-marker – On-screen markers used in the monkey BDM task to indicate the current position of the monkey's bid (red) or the computer's bid (green).

Budget-bar – A rectangular bar whose grey area represents the volume of water that the monkey expects to receive at the end of the trial.

Bid-censorship – The inability of bids to reveal values if those values are beyond the limits of the range of bids that the subject can make.

1 Introduction

This thesis describes the development of a novel method for the elicitation of *subjective values* in monkeys. More generally, we approach this problem with an interest in decision-making, and specifically, with the aim of producing a behavioural method of value elicitation that could be used during a neuronal recording experiment.

Before elaborating on the method developed here, it is first important to establish what is meant by 'subjective value' as well as several other important economic terms. This is made somewhat problematic by the various approaches and definitions that have been offered in the past, often without consensus, and across different fields. As such, the most critical terms are defined here, and are used in a consistent manner throughout this thesis.

Subjective value refers to the value that some subject places upon a reward ***A*** in terms of another reward, ***B***, with the latter reward acting as a *numeraire*. In turn, a numeraire is simply a standard by which value is computed – it is a reward in terms of which the relative values of other rewards may be compared. Thus, the most commonly encountered numeraire is money, and we can achieve this function of money in animal studies by selecting some reward in terms of which the values of other rewards are expressed. The value of a specific magnitude of reward ***A*** can come to be expressed in terms of units of reward ***B***, and the magnitude of ***B*** that is equally valued to that magnitude of ***A*** is entirely subjective.

The values of various rewards in terms of the numeraire allow us to infer and test a subject's preferences across that set of rewards. Rewards with a higher subjective value should be preferred over rewards with a lower subjective value, and this can be tested by examining the subject's pattern of choices when they are presented with two such options. Economists have modelled decision-makers as having an underlying *utility function* which describes their preferences over the various features of rewards. The maximisation of this utility function, given the available options and their costs, amounts to the selection of the most preferred option.

While utility and subjective value are often used interchangeably and should predict the same order of preference over a set of rewards, there do exist some important distinctions between them. We can infer a subject's preferences and their subjective values for different *goods* in terms of a numeraire, however, the subject's preferences over different magnitudes of that numeraire itself are not necessarily linear. The subject also has a utility function that describes their preferences over different magnitudes of the numeraire itself. This may be better understood through an example:

> Take a subject, ***S***, who has a value for reward ***A***, in terms of a numeraire, ***B***, such that the subjective value of ***A*** is equal to 0.75 units of ***B***. Consider that ***S*** also has a value for reward ***C***, such that the subjective value of ***C*** is equal to 1.5 units of ***B***.
>
> While we can conclude from this that ***S*** prefers ***C*** to ***A***, we cannot draw the conclusion that ***C*** is twice as valuable as ***A***, as this assumes linearity in the subject's preferences over different magnitudes of ***B***.
>
> As subjects often show an effect of 'diminishing returns', or concavity of their utility function for a given reward[*], it is likely that the additional utility that ***S*** extracts from each additional unit of ***B*** is decreasing (owing to a negative second differential of the utility function with respect to the magnitude).

In other words, unless the utility of different magnitudes of the numeraire is perfectly linear, then the subjective value determined in terms of that numeraire cannot be substituted for utility without applying some transformation first. In fact, the original application of the Becker-DeGroot-Marschak method (BDM) was to measure the utility of money itself[1] – highlighting the distinction between value, as expressed by a numeraire and utility.

Experimental economists have thus far treated utility like a hidden variable, with subjects acting 'as if' they were maximising some utility function whose form could be inferred from observing the choice behaviour of the subject. Thus, the underlying preferences and utility that describe this have been treated somewhat like a 'black box'. This choice based approach to understanding decision-making "makes clear that the theory of individual decision making need not be based on a process of introspection", and that rather, it "can be given an entirely behavioural foundation"[2].

---

[*] This argument applies equally for the case in which the subject has a convex utility function (i.e. where the second differential of the utility function with respect to the magnitude is positive).

Neuroscience offers some hope that the unobservable black box of utility may in fact come to be observed. Indeed, the reward *prediction error* responses of dopamine neurons of the midbrain have been shown to reflect the behaviourally inferred utility function[3]. Moreover, a reward's utility can be influenced by several factors, including, but not limited to, the timing[4,5] and probability of reward receipt[6,7]. Neuroscientists are beginning to understand how these various reward features are represented neuronally, and how a decision maker integrates these to determine some utility for the rewards in question – with the subjective value of ***A*** in terms of ***B*** reflecting the points at which the ***A*** and ***B*** have the same utility.

The most precise measures of neuronal activity are obtained through direct recordings of neurons in awake and behaving rodents and primates. This thesis introduces the first BDM task especially adapted for, and tested in, monkeys. The BDM allows for more efficient value elicitation than current methods that can be used in monkeys, and, provides another source of cross-species comparisons of decision making as the BDM comes to be increasingly used in human neuroimaging research.

The main data presented here give the monkey's subjective values for different juice-rewards in terms of a water-budget numeraire (Ch. 3.1), and show that monkeys can learn to place bids in a BDM mechanism that reflect their preferences. Utility could have been measured using a paradigm like that of Becker et al. (1964), but this was avoided due to theoretical concerns regarding the validity of BDM values in such a setting (Ch. 2.1). Nevertheless, future experiments could characterise monkeys' utilities for different rewards using the BDM with lotteries as the goods to be bid for.

1.1 - The Becker-DeGroot-Marschak method

The Becker-DeGroot-Marschak method (BDM) is an auction-like mechanism by which an experimenter can determine a participant's true subjective value, in terms of a *budget*, for some *good*. This property of incentivising truthful value revelation is called *incentive compatibility*[8]. A method is incentive compatible if all participants achieve the best possible outcome for themselves by acting in accordance with their true preferences, and this feature of the BDM underlies its popularity in studies of human decision making by experimental economists[9-13], and, more recently, by neuroscientists[14-21].

Although the BDM mechanism takes its name from the authors of a 1964 paper, 'Measuring utility by a single-response sequential method'[1], it is in fact, from the perspective of the subject, strategically equivalent to a Vickrey[22], or second-price sealed-bid, auction, first described and analysed in 1961. In turn, this same mechanism that the Vickrey auction employs actually had its first recorded use in 1797 by Johann Wolfgang von Goethe.

Goethe wanted to sell his epic poem, 'Hermann and Dorothea', but he also wanted to know the value that the publisher placed on his work[23]. So, he devised a mechanism by which he could learn the publisher's value for his poem while selling it at an acceptable price.

First, Goethe set aside a secret 'reserve price' - the minimum amount that he would accept for his poem. He then asked the publisher to provide an offer for the poem. If the offer was more than the reserve price, then the publisher would receive the rights to the poem, paying an amount equal only to the reserve price. On the other hand, if he offered less than the reserve price, then he would lose the auction, pay nothing, and receive nothing. The Vickrey auction and the BDM work in the same way, with the bids of two players[*] in place of Goethe's 'reserve price' and the publisher's 'offer'.

[*] The Vickrey action requires at least two players but can involve more, while the BDM usually has only two 'players' – the experimental subject and a computer bidder.

To understand why the publisher was incentivised to truthfully reveal his value - or willingness to pay (WTP) - for Goethe's poem, we must first understand the optimal strategy under this mechanism[*]:

> Suppose the publisher, $\boldsymbol{A}$, has a value given by $\boldsymbol{X_A}$ and that Goethe's, $\boldsymbol{G}$'s, reserve price is given by $\boldsymbol{X_G}$. By bidding $\boldsymbol{X_A}$, $\boldsymbol{A}$ will win if $\boldsymbol{X_A > X_G}$, and will receive a profit, $\boldsymbol{P}$, such that $\boldsymbol{P = X_A - X_G}$.
>
> Now, assume instead that $\boldsymbol{A}$ places a bid of $\boldsymbol{Y_A}$, such that $\boldsymbol{Y_A < X_A}$. If $\boldsymbol{Y_A > X_G}$ then $\boldsymbol{A}$ still wins with positive profit ($\boldsymbol{P = X_A - X_G}$), however, if $\boldsymbol{X_A > X_G > Y_A}$, then $\boldsymbol{A}$ loses, whereas bidding their true value of $\boldsymbol{X_A}$ would have led to a win with positive profit.
>
> Similarly, a bid of $\boldsymbol{Z_A}$, such that $\boldsymbol{Z_A > X_A}$, can never increase $\boldsymbol{A}$'s profit – it can only introduce the possibility of paying an amount greater than $\boldsymbol{A}$'s value for the good. If $\boldsymbol{Z_A > X_G > X_A}$, then $\boldsymbol{X_G}$ is paid, and, as $\boldsymbol{P = X_A - X_G}$, $\boldsymbol{P}$ will take a negative value, i.e. $\boldsymbol{A}$ incurs a loss.
>
> Bidding one's true value is therefore never worse, and sometimes better, than bidding any other value.

The above mechanism holds if $\boldsymbol{G}$ is a computer bidder, randomly selecting a bid, $\boldsymbol{X_G}$ - both the optimal strategy and the underlying logic are the same. Bidding one's true value is a *weakly dominant strategy* in such sealed-bid, second-price, mechanisms. In fact, Goethe's optimal strategy would also have been to select a reserve price that was equal to his value for the poem, as optimal selling – or willingness to accept (WTA) - prices should also be equal to the seller's value for the good[†]. Similarly, the original BDM experiment involved subjects selling a reward - though buying versions of the task are now more commonly used.

A few assumptions are necessary for this logic to hold, however, and most of these are captured by the axioms that describe an expected utility (EU) maximising subject[24] (see Appendix 1) – the possible effects of violating these axioms are

[*] This is described formally in Appendix 1 – Optimal BDM strategy.

[†] Buying and selling prices in the BDM should be equivalent, but see Ch. 2.1, for a short discussion of observed willingness to pay (WTP) – willingness to accept (WTA) disparities and the endowment effect.

addressed at a later point (Ch. 2.1). For now, the only other assumption of concern is that the bidders' values are private. This means that they do not know each other's values, and, that such knowledge would not alter the value of the object for either bidder - a reasonable assumption when the object's value is derived from its consumption alone[25].

Unfortunately for Goethe, it was precisely this latter assumption that would be violated. Goethe had left his reserve price in a sealed note with a friend who would go on to reveal this information to the publisher, who then submitted a bid at exactly that amount. Although Goethe's revenue was not affected by this – all bids above the reserve price would have paid that same amount – he did lose out in another way: he learnt nothing of the publisher's subjective valuation of his work, apart from the fact that it was at least equal to his reserve price.

In its neuroimaging applications, it is this revelation of subjective value information for which the BDM is employed. To identify those areas and mechanisms of the brain that can make decisions between options that might differ even in their fundamental objective features (e.g. choosing between food and a film), it has been supposed that the different options are assigned a subjective measure of their desirability, or a utility, that allows for their comparison on a common scale[6,26,27]. Thus, the option with the highest net utility can be chosen - having accounted for any costs incurred in acquiring it.

As such, neuroscientists have sought to distinguish between the objective measures of a reward, such as its magnitude, and the utility that the participant places upon it - though subjective value measures in terms of a numeraire are often used in place of utility itself. This has allowed for identification of those brain regions[28,29] and neurons[30,31] that are thought to be responsible for representing subjective values and mediating choice. Consequently, a method that allows for the truthful revelation of subjective values on a trial-by-trial basis is a powerful tool for the neuroscientist with an interest in decision-making.

In fact, with regards to the BDM, it was the property of trial-by-trial value revelation that had initially motivated its design, with the authors presenting the mechanism as a 'single-response' method.

After all, subjective values could already be inferred in an incentive compatible manner by utilising a simple binary-choice (BC) method. On each trial of a typical BC task the subject is offered some fixed amount of reward ***A*** alongside some variable amount of reward ***B***, and chooses one of these two options. Each pair of options is repeatedly presented to the subject, such that for each magnitude of ***B*** (i.e. for each ***A-B*** pair), the probability of choosing ***B*** is estimated from the proportion of trials on which ***B*** was chosen. The amount of ***B*** for which ***A*** and ***B*** would be expected to be chosen with equal probability is called the *indifference point*, and is an estimate of the subjective value of ***A*** in terms of ***B***.

The BC method depends upon discrete choices and a rationalisation of probabilistic choices, informed by the psychophysics literature, which supposes that “a gradation of preference is the rule when persons locate themselves on “physical” continua”[32]. Becker et al. (1964) identified the key weakness of such discrete choice methodologies as lying in their assumption that “the subject’s probabilities of choice remain constant throughout the many times that he is choosing from the same available set of actions”. The BC task requires many trials over which choices are pooled and during which preferences and values are assumed to have remained unchanged. When one considers temporally associated effects, such as satiety, it is clear why such a method might be problematic.

This problem is compounded by the fact that values must be inferred from the subject’s behaviour in the BC task by fitting a logistic function[*] to their choice proportions for the variable reward. The best fitting logit curve is sensitive to noise and many trials must therefore be tested to diminish the influence of erroneous decisions (fig.1.1). Moreover, there is rarely any data at the point of indifference itself – more often, this is simply an inference based upon an interpolation between the choice probabilities that are above and below 0.5.

In the BC task, the subject’s inferred value is a low temporal resolution estimate based upon a model of the subject’s choices, while in the BDM the subject makes a direct trial-by-trial report of their subjective value, or WTP. It is not difficult to see why such a method would gain popularity in neuroimaging studies of decision making.

---

[*] The statistical methods that are used to identify the indifference point in a binary choice task are described in Appendix 3 – Logistic Regression**.**

With a limited number of trials in costly fMRI experiments, the 'single-response' property of the BDM is not just a theoretical, but also a practical, consideration.

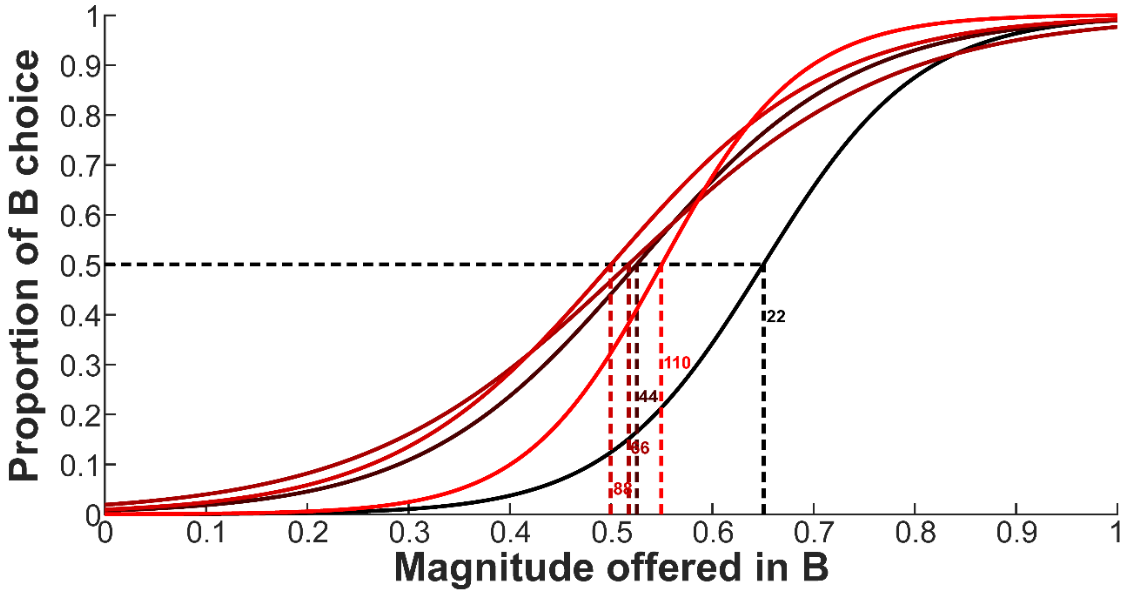


Fig. 1.1) A simulated decision-maker in binary choice tasks with different numbers of trials. The best fitting logit model is shown for different numbers of trials, which are presented to the left of the relevant dotted line, indicating the estimated indifference point. The simulated decision-maker had an underlying value of 0.55 units of ***B*** for another reward, ***A***, with some noise in the representation of the value at the time of decision-making*. The addition of more trials generally improves the accuracy of the value estimate.

Plassman et al. (2007) conducted the first BDM experiment on subjects in an MRI scanner[14], showing that the orbitofrontal cortex encodes the subject's value for different food rewards. Moreover, they emphasised the importance of the 'single-response' property of the BDM, with the fact that it gives "a measure of the WTP computed by the brain for every bidder and item at the time of decision making" forming part of their rationale for using the method. Notably, Plassman et al. (2007) also went on to describe the BDM as a decision-making task that is "significantly more abstract and complex than those that can be studied in monkey experiments". This thesis hopes to challenge that view.

* Simulation methods are detailed in Appendix 4.

A BDM task for use in monkeys:

Currently, the only available method for the identification of subjective values in monkey experiments is the BC task[30]. Unlike humans, who can perform well having been told that only a minority of their trials will be actualised, monkeys require regular receipt of reward - usually on a trial-by-trial basis - to learn, remain motivated, and perform meaningfully in a task. Monkeys therefore become satiated as they perform the BC task, limiting the number of rewards for which a value can be inferred, and, potentially reducing the accuracy of the inferred subjective value (fig. 1.2).

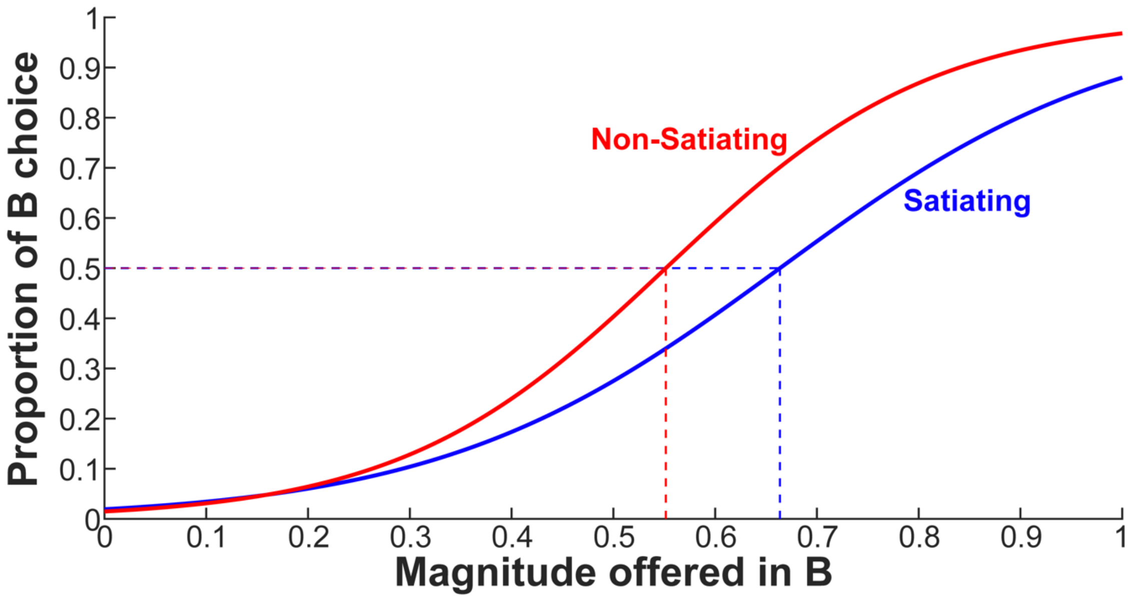


Fig. 1.2) The best fitting logit models are shown for simulated non-satiating and satiating decision-makers over 55 trials (5 repetitions of each pairwise ***A***-***B*** comparison). Both start with a value of 0.55 units of ***B*** for another reward, ***A***. The satiating decision maker updates their value of ***A*** based upon the amount of ***A*** and ***B*** that have been consumed (consuming more of ***B*** increases the relative value of ***A***, as it is assumed that sensory specific satiety[33] for ***B*** reduces the subjective value of ***B***, while consuming more of ***A*** reduces the value of ***A*** for the same reason – see Appendix 4). The satiating decision maker finished the task with a value of 0.61 units of ***B*** for ***A***, with a mean value of 0.58 over the course of the session, but the value estimated by the logit model was 0.66. On the other hand, the estimate for the non-satiating decision maker was correct, at 0.55.

Given these additional limitations, it could be argued that the BDM may be even more useful for monkey experiments, particularly those using neurophysiological recording, where the experimenter has a limited number of trials with which to test an isolated neuron. Studies of decision making in monkeys would benefit greatly from the larger number of rewards and scenarios with which the properties of a value coding neuron could be tested – if values can be estimated quickly, then this leaves more time for testing alternative conditions and controls.

In addition to this, the high temporal resolution of value estimation in the BDM could allow for more precise measurement of the effects of a given manipulation. For example, the BDM could be used to quickly determine how a monkey’s value (and neuronal coding of value) changes following receipt of a large bolus of reward, or after application of a pharmacological lesion to a decision-related brain region.

However, the advantage of the BDM method does not lie only in its ‘single response’ property. The BDM elicits these subjective values in a rich decision environment. That is, unlike most BC tasks, the BDM elicits probabilistic expectations, gains, losses and values in a single trial – all of which are variables that contribute to an animal’s perception and decision making in the trial, and for which there presumably are some underlying neural correlates.

Of course, the fact that using a given method would be advantageous in some species if they could perform it does not mean that they can. Indeed, the most obvious problem in developing this method for monkeys lies in the fact that they cannot be told what the optimal strategy is. While human subjects are usually instructed in the optimal strategy, monkeys must learn the BDM through trial and error.

Moreover, the simplest BDM task has four primary elements, without which the task cannot feasibly be learnt: a budget with which to place bids, a representation of the reward to be bid for, a subject bid, and, a computer bid. These elements are all dynamic: their significance must be learnt and their values tracked if the monkey is to make reasonable bids in the task. To further complicate matters, there are many different possible actions - as opposed to the two possible actions in a BC task – and the reward outcomes depend on complicated probabilistic contingencies between action and outcome, with a high degree of variability in the payoffs associated with each possible bid.

Despite these complications, the problem that the monkey faces nevertheless seems to be fundamentally tractable. The complicated probabilistic contingencies between action and outcome only limit the rate at which learning occurs. The high degree of variability in possible outcomes, due to the random generation of computer bids over the range of possible values, should increase the number of trials that are needed to learn the expected payoff for a given bid. However, the contingencies themselves don't necessarily need to be learnt. In theory, a subject learning by reinforcement[34] could come to make rational bids in the BDM simply by tracking the mean payoff of different bids. If we consider the BDM to be analogous to a multi-armed[35] *bandit-task*, whereby the monkey has some number of possible bids, each with a different distribution of rewards, then the limitations on the monkey's ability to perform the task appear more quantitative than qualitative, as we know that monkeys can and do perform rationally in simple bandit-tasks[36].

Unfortunately, the payoff function for the BDM is shallow (Ch. 2.1), meaning that the difference in mean payoffs for different bids is small relative to the optimal payoff – the subject makes only small losses relative to the optimal payoff for sub-optimal bids. As the payoff gradient drives learning, the shallow payoff gradient of the BDM acts as an impediment to learning, though bids should still be driven towards the optimum. One advantage of using monkeys over human subjects is the vastly larger number of trials and sessions with which they can be trained, and so this problem may be overcome by using a longer training period, which is quite normal for monkey studies of decision making.

As such, we hypothesised that monkeys could learn to make bids in a BDM task with sufficient training. In this thesis, we present evidence in support of this hypothesis, and analyse the behaviour of two monkeys who were taught to place bids in a BDM task (Ch. 3.1). The monkeys learnt to use a joystick to report a bid in terms of a budget of water for different volumes of a juice-reward. They were also trained in a binary choice task (Ch. 2.3) that was adapted[*] to be an analogue of the BDM. Thus,

[*] Chapter 5, Section 3, for a description of the differences between a typical binary choice task and the binary-choice-bundle (BCb) task, and a discussion of the rationale for using the latter.

we could compare the monkey's values for the different rewards using two independent incentive compatible mechanisms, allowing us to probe the accuracy and usefulness of the monkeys' behaviour in the BDM.

The BDM is a more complicated paradigm than the BC task, and this was evident in the fact that it took us approximately 3 years to achieve a satisfactory level of performance in the two monkeys, while both learnt to perform consistently in a BC task over the course of just six months. However, as the field of decision-neuroscience progresses, more complex questions will present themselves and these may require more complex tools with which to probe animal behaviour and its underlying neural correlates. The BDM task for monkeys provides a new, richer, and more efficient, value elicitation mechanism – a new tool for the behavioural and neurophysiological study of decision making in monkeys.

1.2 - Organisation of the thesis

The final version of the BDM task for monkeys is described in Chapter 2, after theoretical considerations with regards to the BDM and the findings of previous BDM studies are considered, all of which helped to form the rationale that guided the task design. The results and analysis of the behaviour of two monkeys in that BDM task is then presented in Chapter 3.

Chapter 4 goes on to detail the training steps that were necessary to teach the monkeys the various components of the BDM task, as well as how to use the joystick to place bids, ending with the results of a preliminary version of the BDM task in the first monkey.

Several exploratory studies into different versions of the BDM, as well as the development of the supplementary binary-choice task, are described in Chapter 5 – these intermediate experiments led to the development of the final version of the task presented in Chapters 2 and 3.

Finally, an experimental design for use in neurophysiological experiments is outlined in Chapter 6. In particular, that experiment would make use of the BDM to quantify expected costs and their associated prediction errors while simultaneously collecting reports of the monkeys' subjective values.

The same methods and theory are used and referred to repeatedly for various versions of the BDM task throughout this thesis. Therefore, to avoid unnecessary repetition, these are contained in the appendices, with reference made to them where relevant. These appendixes include: BDM theory (Appendix 1), experimental methods (Appendix 2), statistical methods (Appendix 3), and, simulation methods (Appendix 4). For ease of reference, different versions of the BDM task are assigned names when they are first introduced, and these are displayed in square brackets for purposes of clarity.

2
BDM task design

In this chapter, we describe the design of the final version of the BDM task for monkeys. First, we look back at some past experiments that have utilised the BDM method in both behavioural economics and neuroimaging applications (Ch. 2.1), and consider several theoretical issues that guided various aspects of the task design. We determined those features of the BDM that can be adapted for use in monkeys and those that need to be changed, with the key differences arising from the need to aid learning of the optimal strategy in the absence of instruction or explanation of the underlying reasoning.

Importantly, given the novel, untested, and unproven status of this task in monkey experiments, we wanted to avoid using a version of the task for which there are considerable theoretical concerns - supported by experimental data - regarding its incentive compatibility. We therefore sought to design a version of the BDM that would rely upon as few assumptions as possible regarding the nature of the monkey's preferences and their supposed utility function.

After explaining the rationale that motivated our BDM task design, we move on to describe the task itself (Ch. 2.2). Finally, the design of a supplementary binary-choice-bundle (BCb) task is also presented (Ch. 2.3). The BCb task is an especially adapted binary-choice task that was necessary to allow for assessment of the reward values estimated by the BDM method. An established incentive compatible method of value elicitation was required as a benchmark with which to compare the results of this novel BDM in monkeys.

It's important to note here that the BDM task described in this chapter was the result of extensive testing with different versions of the BDM over approximately three years, and this process is described in Chapter 5. Moreover, the steps required to teach a monkey the various stimuli that signify the different task elements, as well as how to place bids in the first place, are described in Chapter 4. The technical experimental methods, relating to the display of stimuli on the monitor, the monkey's interface with the task using the joystick, and the delivery of rewards, all remain unchanged between various versions of the BDM task. Therefore, to avoid repetition, the technical aspects relating to experimental methods are described together in Appendix 2 – Experimental Methods.

2.1 – BDM experiments and theoretical considerations

The BDM was first proposed as a method for measuring the utility of money. The authors used a sequential method in which two subjects were offered a *lottery* – a set of possible outcomes, each with a stated probability (in this case over different monetary amounts) and had to state a selling price for it - or a willingness to accept (WTA)*, referring to the price at which a sale would be accepted. A subject's optimal price for a lottery is said to be their *certainty equivalent (CE)* for that good: the amount of money, to be received with certainty, such that the subject is indifferent between it and the lottery.

The certainty equivalents can be used to determine a utility function for the subject through an iterative process, or *fractile method*†. For an expected utility (EU) maximiser (Appendix 1), the CE for a lottery between two outcomes with probabilities $\boldsymbol{p}$ and $\boldsymbol{1-p}$, is taken as the magnitude of reward at which the utility equals the probability-weighted utilities of the two lottery outcomes. Thus, where ***u*** is the subject's utility function for a given reward, and ***A*** and ***B*** are two different magnitudes of the reward, the utility at the certainty equivalent (CE) is given by:

$$\boldsymbol{u(CE) = pu(A) + (1-p)u(B)}$$

As utility is an abstract scale, any magnitude can be determined as having a utility of 1, and we define an outcome with magnitude zero as having a utility of zero. Substituting ***u(B) = 0*** and ***u(A) = 1***, where ***B = 0*** and ***A > B***, we find that:

$$\boldsymbol{u(CE) = p}$$

Thus, a typical fractile method proceeds by setting ***p = 0.5***, and finding the CE, ***C***, between ***A*** and ***B***, where ***u(B) = 0*** and ***u(A) = 1***. We now know that this magnitude ***C*** has a utility equal to ***p***, such that:

$$\boldsymbol{u(C) = 0.5}$$

* Whether selling or buying, the optimal BDM strategy is the same.
† The fractile method can also be used, more slowly, in the context of a BC task. The CE is taken as the indifference point that is estimated for the lottery over a reward, ***R***, by means of logistic regression over the choices between the lottery and different certain amounts ***R***.

This gives the first derived point, ***C***, on the utility function (fig. 2.1a) and can be used to derive the next. The certainty equivalent, ***D***, for a lottery between ***C*** and ***A*** with the same value of ***p***, 0.5, is the magnitude of reward with a utility of 0.75 (fig. 2.1b):

$$\boldsymbol{u(D) = pu(C) + (1-p)u(A)}$$

$$\boldsymbol{u(D) = 0.5p + (1-p) = 0.25 + 0.5}$$

$$\boldsymbol{u(D) = 0.75}$$

This process can be repeated for lotteries offering different potential magnitudes as outcomes, to map the relationship between reward magnitude and utility*.

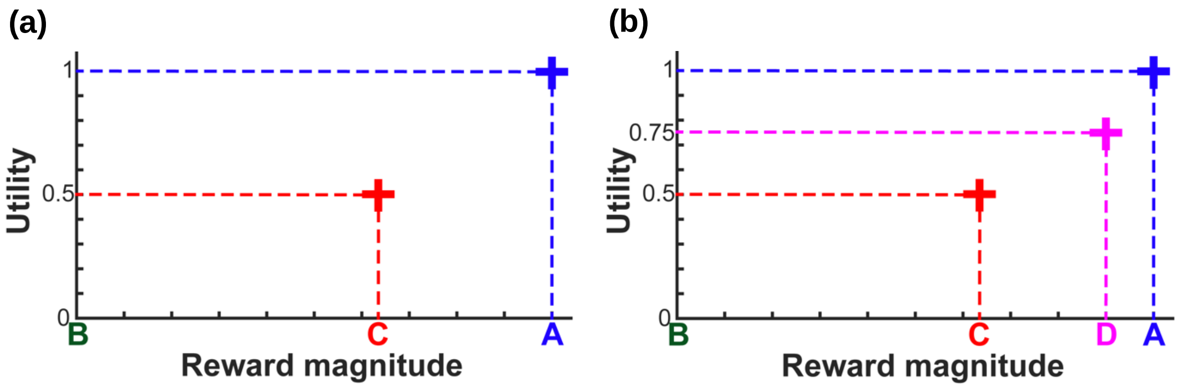


Fig. 2.1a) When the magnitude ***A*** is set to have a utility of 1 and ***B*** has a utility of 0, then a lottery with ***p = 0.5***, will have a CE, ***C***, with a utility of 0.5. **b)** Similarly, a lottery offering ***A*** or ***C*** at ***p = 0.5*** has a CE, ***D***, with a utility of 0.75.

It can also be shown that different lotteries should have the same utility, and therefore the same CEs, under the assumptions of EU theory. Take, for example, the lottery between ***D*** and ***E***, such that ***u(E) = 0.25****,* with CE, ***F***:

$$\boldsymbol{u(F) = pu(D) + (1-p)u(E)}$$

$$\boldsymbol{u(F) = 0.75p + 0.25(1-p) = 0.375 + 0.125}$$

$$\boldsymbol{u(F) = 0.5 = u(C)}$$

* Different probabilities can also be used but we have used *p = 0.5* here for simplicity.

In this case, ***F*** must be equal to ***C***. Becker et al. (1964) used this fact to test their utility model (an EU model), and rejected it due to the inconsistency of subjects' WTA reports for lotteries that should have had the same utilities and certainty equivalents. However, they did note that "as the subjects became more familiar with the task their experience appeared to lead to more consistent behavior and less deviation from the results specified by the utility model".

Indeed, later experiments using the BDM mechanism would utilise more extensive training and instruction of subjects, as well as using *induced value* bidding procedures - in which the optimal bid is known to the experimenter[37,38] - to aid in more precise measurement of the subjects' deviations from optimality.

Nevertheless, researchers have continued to find inconsistencies in the bids of their subjects when using a lottery as the good in the BDM. One possible explanation for this, including the results of the original BDM paper, lies in the possibility that the BDM is not incentive compatible for lotteries[39,40]. If subjects violate the independence axiom (Appendix 1), and do not weight probabilities in a linear manner, then the value of the compound lottery that they face in the BDM, when the reward to be bid for is itself a lottery, need not be the same as the lottery that is being bid for – that is, the compound lottery cannot necessarily be reduced to a simple lottery (Appendix 1).

Further concerns regarding the use of lotteries in the BDM relate to the effects of *probability distortion*, which has been observed in both human[41] and monkey[42] behavioural experiments. Probability distortion describes a subjective non-linear weighting of probabilities, and typically leads to the overweighting of low probability outcomes and the underweighting of high probability outcomes. Becker et al. (1964) acknowledged the possibility that the subjects' "personal probabilities" for the outcomes did not match their objective probabilities, and therefore used only two different probabilities, ½ and ¾, which they assumed would be well understood by subjects. Nevertheless, it is possible that the different weighting of these two probabilities still contributed to the inconsistent valuations that they observed for lotteries that should have otherwise had the same expected utility.

In addition to the inconsistencies associated with the use of lotteries, other deviations from theoretical expectations in the BDM have also been observed. Perhaps the most well-known of these is the disparity in bids between willingness to pay (WTP) and willingness to accept (WTA) versions of the BDM[43], with WTA being found to be systematically higher than WTP for the same objects - a finding which has also been replicated in neuroimaging experiments that utilised the BDM[15]. The *endowment effect*[44] – the notion that subjects place a higher value on objects they own - has been proposed as an explanation for this difference.

From a theoretical standpoint, however, it has been pointed out that *income and substitution effects* could also explain the WTP-WTA disparity[45], and indeed, the work of Shogren et al. (1994) has shown that the availability of close substitutes does reduce the WTP and WTA disparity[46]. Of greater relevance to the BDM mechanism itself, however, is the finding that whilst repeated experience in other auction mechanisms can lead to the convergence of WTP-WTA measures, this is not the case for the BDM[47]. Furthermore, it has been shown that varying the experimental procedure by which WTP and WTA are measured can lead to the WTP-WTA gap being “turned off and on”[48]. These results suggest that the endowment effect is not a fundamental feature of preferences but rather an artefact of the BDM itself, which the authors suggest is due to systematic misconceptions that subjects hold with regards to the mechanism.

Indeed, if such effects are a result of subject misconceptions or the effects of an interaction of the task with an exogenous market, then it is possible that some of these inconsistencies may not be found in the monkey subject. For example, it has also been suggested that inconsistencies in the BDM may be due to the fact that subjects are not “interested in the type of objects used in the experiments”[49], but such a concern is less likely to hold in the case of a monkey, for whom the task represents their main source of liquid intake on training days and where there is no apparent external context that could interfere with their valuations.

More pertinent to the adaptation of the BDM for use in a monkey experiment are the concerns regarding the low payoff gradient that the subject faces in the BDM (fig. 2.2), especially given the fact that the monkey cannot be instructed with regards to the optimal strategy or the reasoning behind it, but must rather search for it. As a greater difference in the expected payoffs of different options can drive faster learning, this shallow payoff gradient may constitute a significant obstacle in training monkeys to perform the task. Moreover, even in the case of human subjects who understand the optimal strategy, the shallow payoff gradient over the range of possible bids could make the costs of deviation from optimality small when compared to the cognitive costs of determining and making an optimal bid – as has been argued in the case of the *first-price auction*[50].

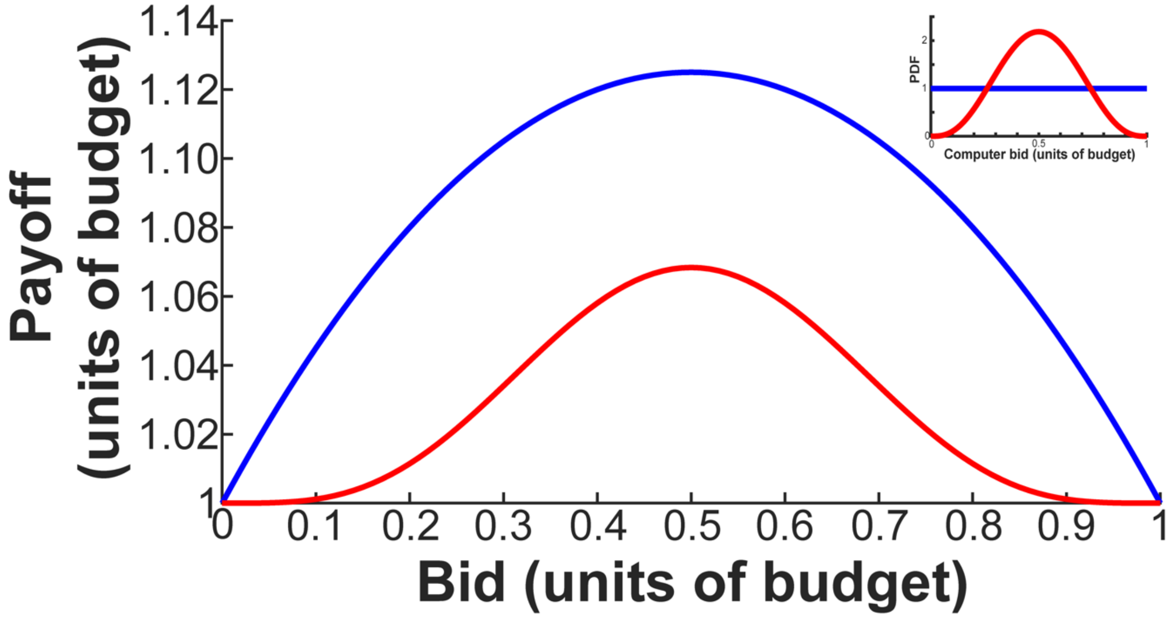


Fig. 2.2) 'Steep' and 'shallow' (or 'peaked' and 'flat') payoff schedules are shown for different bids of a simulated subject. The payoff schedules are shown for a simulated subject bidding for an object with a value of 0.5 against a computer bidder drawing their bids from a 'steep' Beta (4,4)* distribution (red) and a 'flat' uniform distribution (blue). The distributions themselves are shown inset. The maximum possible payoff is found where bid = value, which is 0.5 in both cases – i.e. the steepness of the payoff schedule doesn't influence the optimal strategy. See Appendix 4 for simulation methods.

This payoff gradient can also be expressed in terms of the 'expected cost of misbehaviour' (ECM)[51], which describes the difference between the expected profit of an optimal bid (i.e. where the bid is equal to the subject's value for the good) and the expected profit of some other sub-optimal bid (fig. 2.3)†.

---

* The Beta distribution is parametrised by two positive shape parameters, α and β. The Beta distributions used here are described using the following format: (α, β).
† The ECM is formulated in terms of the differences in profit, and these are equivalent to the differences in payoffs. See Appendix 1 for details of how the ECM is calculated.

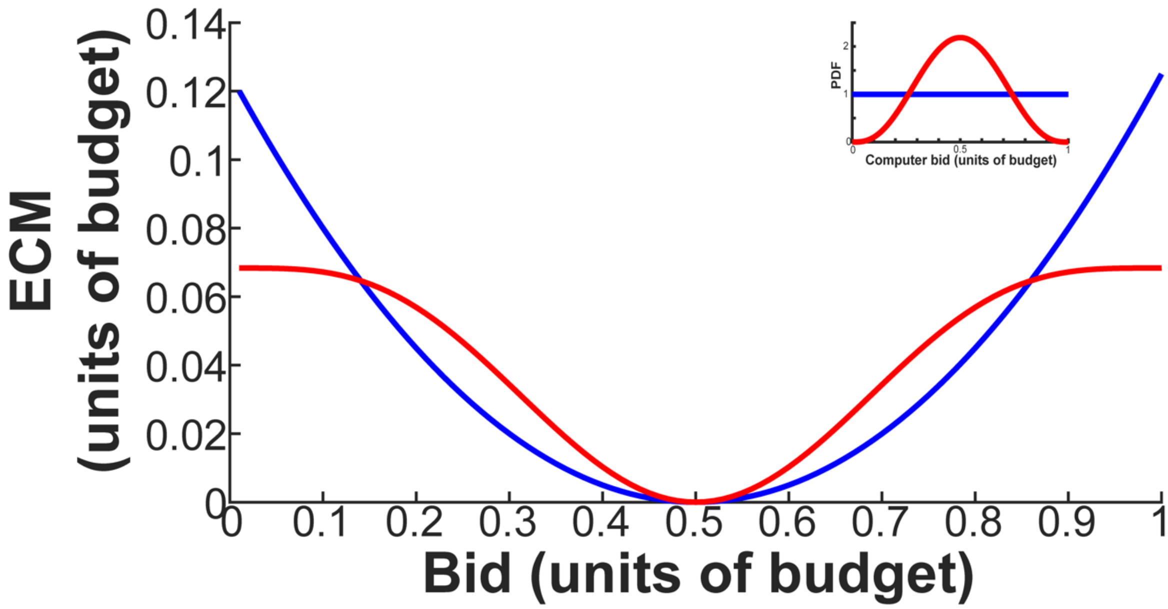


Fig. 2.3) The expected cost of misbehaviour (ECM) varies depending on the distance of the bid from the optimum, as well as varying with the computer-bid distribution that is used. Note that the optimal strategy does not change: the cost is always zero when the subject's bid is equal to their value for the good (in this case value = 0.5). The ECMs are given for two different computer-bid distributions, a Beta (4,4) distribution (red) and a uniform distribution (blue) – these distributions are shown inset.

Note the relatively low costs of deviating from optimality in the BDM - take for example the simulated bidder shown in figure 2.3, with a value of 0.5 facing a uniform distribution: for a bid of 1, 100% more than the subject's value, the expected cost of such a bid is only 0.12 units, or ~11% of the optimal expected payoff (in this case 1.12 units of the budget – see fig. 2.2).

This shallow payoff gradient provides a potential source of inconsistency in a subject's BDM bids. In addition to this, motor noise, perceptual noise, and a possible element of stochasticity in the subject's underlying utility for the reward[52], could all contribute to variability in the subject's bids for the same reward, or, for different rewards that might be expected to have the same utility. Therefore, although the BDM does provide a 'single-response' paradigm, collecting more bids for each reward still plays an important role in increasing the certainty regarding the subject's underlying value for the good.

The effects of using different payoff schedules in the BDM were investigated in several experiments by Irwin et al. (1998), who studied the effects of manipulating the amount of instruction that subjects received[37], the payoff schedule that was used, and the feedback that they provided on each trial[37]. In their first experiment, they simply demonstrated the incentive compatibility of the BDM in an induced value setting[*].

Irwin et al.'s (1998) second experiment investigated the effects of using different payoff schedules (flat vs. peaked) in conjunction with providing subjects with different levels of information (full vs. minimal) before testing their BDM performance. Full information subjects were told the value of the object being bid for, the probability distribution from which the computer bid was drawn, the computer's realised bid, and whether they had won or lost each round of the auction. In contrast, minimal information subjects were only told that there was an optimal bid between the minimum and maximum and that their task was to find this bid – they had no information with regards to the process by which the price was determined, or even the price that they had to pay on a given trial: they were only told their initial balance, payoffs and their final balance. In effect, these minimal information subjects had to search for the optimal strategy.

The findings of this second experiment were clear, the bids of subjects in the full information condition were close to optimal and became more accurate with each additional trial, regardless of the payoff schedule that was used. The bids of subjects in the minimal information condition were also unaffected by the payoff schedule, but their performance deteriorated as the experiment went on, and their bids were much further from the optimum (fig. 2.4). These findings led the authors to conclude that the randomness in the feedback provided by the BDM "defeats any attempt to search systematically for the optimal bid".

A third experiment investigated the effects of trial-by-trial feedback for subjects in the minimal information condition. Again, subjects were not told how the bidding mechanism worked, but their balance was changed on each trial according to the expected values of their bids. With this change in feedback a difference in bidding behaviour was now observed between minimal information subjects facing steep and shallow payoff schedules.

---

[*] Interestingly, they also found no evidence of a WTA-WTP disparity in their subjects.

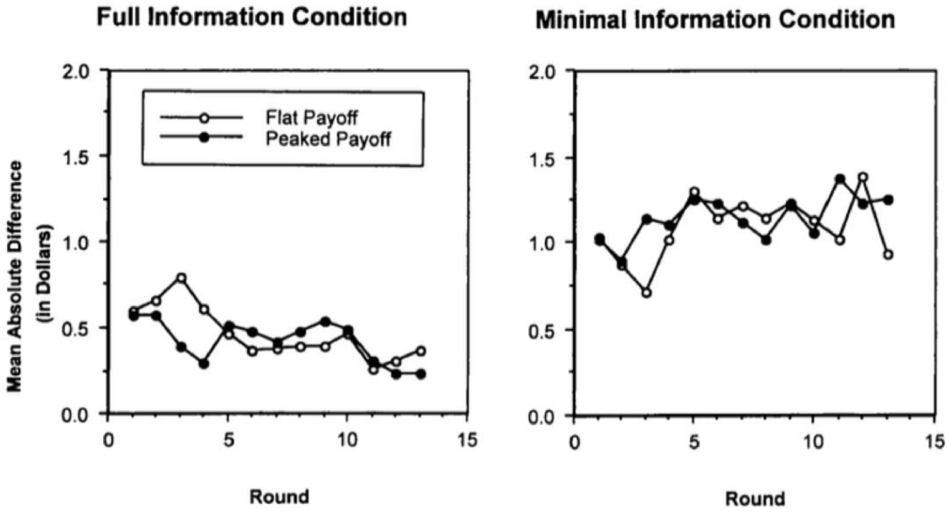


Fig. 2.4) The mean absolute difference between the subject's bids and the optimal bids for full and minimal information subjects under both 'flat' and 'peaked' payoff schedules in Irwin et al.'s (1998) second experiment - this figure is adapted from figure 6 of that paper, [37].

The crucial finding of this experiment was that in the low information condition manipulation of the payoff schedule alone was not sufficient to make the subjects' bids more efficient – payoffs had to be changed alongside clear trial-by-trial feedback to improve the consistency and optimality of the subjects' bids. Thus, the authors concluded that "to be useful, feedback must provide the subject with the link between the decision and the payoff and must be provided before the next decision is made".

These findings played a significant role in guiding the design of our BDM task for monkeys. Even the most basic instructions cannot be provided for our monkey subjects – their 'instructions' are the apparatus, symbols, actions and payoffs of the task itself. Our monkey subjects are therefore most analogous to the minimal information condition subjects of Irwin et al.'s (1998) second and third experiments, and it does not seem unreasonable to suppose that if these measures are necessary for improving the bids of human subjects - who have at least some familiarity with the concepts involved in bidding and auctions - then they are likely to be vital in supporting learning of the BDM and its optimal strategy in monkeys.

Thus, while we would have preferred to mimic human BDM experiments by resolving only a fraction of BDM trials – limiting the effects of satiety – we had to display and realise the outcome of each BDM trial before progressing to the next, providing the monkey with a direct link between the decision and the payoff.

Another important difference between our BDM study and most of those tested in human subjects is that we have avoided the use of lotteries, or risky rewards, due to the theoretical concerns that some have expressed regarding violations of the independence axiom (see above). Thus, due to the exploratory nature of implementing a BDM task in monkeys, we proceeded in a similar vein to Wertenbroch and Skiera (2002), who used the BDM to elicit WTPs for certain, non-random, goods only[53]. This prevented us from measuring utility directly using a fractile method, instead, we sought to first establish meaningful performance of monkeys by eliciting their subjective values for simple, riskless, rewards (Ch. 3.1).

Nevertheless, there do still exist some challenges to the incentive compatibility of the BDM even for non-random goods[54]: if the subject's utility function has a form like that proposed by Machina[55] - where a local utility function describes preferences in terms of the probability distribution that the subject faces – then the values that the BDM elicits are not necessarily incentive compatible.

Therefore, we cannot but rely on some assumptions regarding the nature of the monkey's preferences, though notably, our underlying assumptions remain weaker than those of previous studies using risky reward objects. We assume that the monkey's preferences are invariant to the probability distribution of computer bids; that the values are private (this is trivially satisfied in our task); and, that the monkey's preferences are complete, transitive, and continuous (Appendix 1).

To summarise, an overview of theoretical concerns regarding the BDM has led us to design a task in which each trial is realised before the next commences, and, to avoid the use of risky reward objects, at least in our first demonstration of monkey behaviour in the BDM.

2.2 – Monkey BDM task design

The focus when developing this task for monkeys was on portraying the various elements of the task simply and with clearly associated contingencies and reward outcomes. The BDM task has four fundamental elements: a good to be bid for, a budget with which to bid, the subject's bid, and the computer's bid. Interactions of these elements can explain the mechanism and can be used to predict the outcomes. Two monkeys, Ulysses and Vicer, were taught to associate each of these task elements with a visual stimulus and used a joystick to move a *bid-marker* within a rectangular *budget-bar* that represented their budget for that trial – the resting position of the monkey's bid-marker signalled their bid for that trial.

The BDM task design is described in this section, while more technical experimental methods can be found in Appendix 2. First, the task stimuli and the way in which monkeys place bids using a joystick are described. Following this, the structure of each trial and the time course of stimulus changes and reward receipt are outlined. The section ends with a discussion of the overall organisation of testing sessions that were used to generate the data that are presented in Chapter 3.

Task stimuli:

The monkeys were first trained[*] to associate several stimuli with the different basic elements of the task (fig. 2.5). We used 0.15, 0.3, 0.45, 0.6, and 0.75ml mango juice-rewards as five different goods that the monkey could bid for. These volumes of juice-reward were selected based on their water-budget values as inferred from an analogous binary-choice-bundle (BCb) task and chosen such that the 5 different rewards covered the range of possible bids (Ch. 3.2). The monkeys were taught to associate each of these volumes of mango juice-reward with a different fractal image (reward-fractal) in a preceding Pavlovian stimulus learning task. On each trial the monkey was also presented with a 1.2ml budget of water that could be used within that trial and was represented by a grey rectangle, or 'budget-bar'. The monkey

[*] See Chapter 4, Section 1, for details of the Pavlovian stimulus learning tasks that allowed the monkey to acquire these associations.

could use the budget make bids, ‘spending’ a certain proportion of it when making successful bids and receiving the remaining water at the end of each trial. When making a losing bid the monkey would receive the entirety of the initial 1.2ml budget, but not the juice-reward. The volume of available budget at any given time was indicated by the grey area of the budget-bar, with any budget that had been paid in the BDM being shown by the black occluded area of the budget-bar. For example, if the monkey had bid 1ml and the computer had bid 0.9ml of water-budget, then the monkey would win the auction, pay 0.9ml of water, and receive the juice-reward. In this case, the bottom three quarters of the budget-bar area would be occluded to represent the 0.9ml cost (¾ of the total 1.2ml water-budget), and the remaining un-occluded 0.3ml of water-budget would be received at the end of the trial.

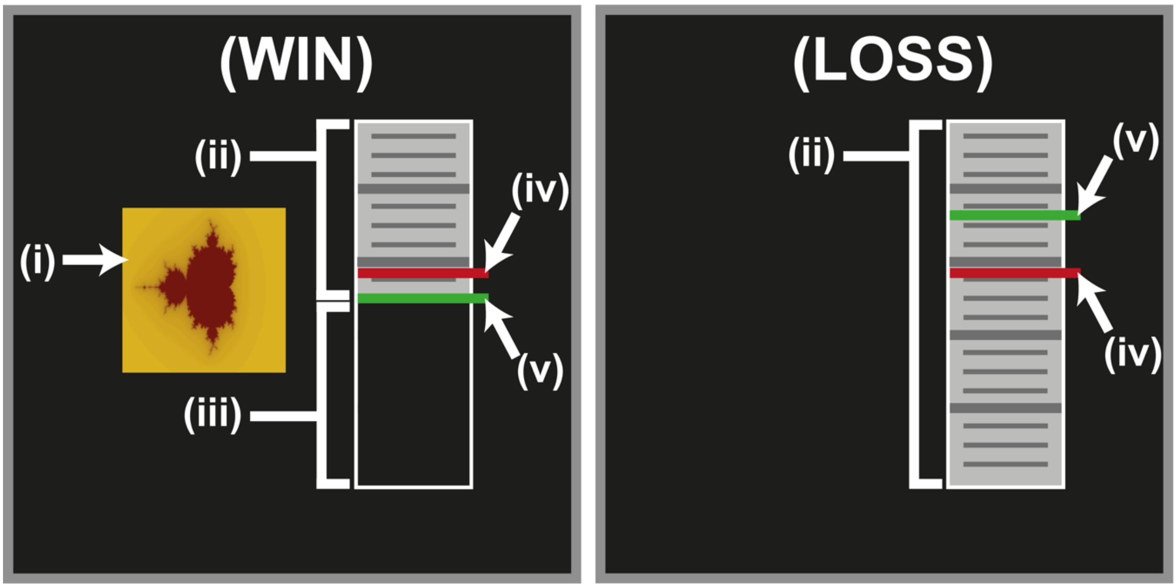


Fig. 2.5) The key associations that the monkeys were taught are shown in the context of the display of outcomes for both win (left) and loss (right) trials. The good that the monkey bids for on each trial is indicated by a fractal **(i)**, and monkeys were taught to associate 5 different fractals with 5 distinct volumes of a mango juice-reward. The remaining volume of water-budget **(ii)** is indicated by the area of a grey rectangular ‘budget-bar’, with the amount of water-budget that has been paid being signified by the occluded portion of the budget-bar **(iii)**. The monkey’s bid is indicated by a red ‘bid-marker’ **(iv)** that can be moved up and down the budget-bar using a joystick. After the monkey has made their bid, a green ‘bid-marker’ **(v)** appears to signal the computer’s bid. The height of the bid-markers indicates the magnitude of the bids as a proportion of the total budget-bar volume. See figure 2.6 for a description of how these stimuli change over the course of a single BDM trial.

The total starting area of the budget-bar was outlined by a white border which was never occluded, thus the monkeys could always see the amount of water-budget that had been paid and the amount that was remaining relative to the initial starting budget – emphasising the relationship between the computer-bid position and the costs on each trial.

Maintaining a contrast between the budget and the good was desirable to assist the learning of different associations for each of these rewards. As such, we had considered using a food-budget, but, as the monkeys were only liquid restricted throughout the week this would have led to weaker control of access to the budget, which in turn could lead to greater variability in bids on different days.

Both the water-budget and the juice-reward were delivered at a liquid spout, and this was mediated by the opening and closing of a computer controlled solenoid (Appendix 2).

In addition to the reward indicating stimuli we also used two different ‘bid-markers’. The monkey’s bid was indicated by a red bar which could be moved up and down within the budget-bar, and the monkey controlled the movement of this red bid-marker by moving a joystick. Forward movements of the joystick moved the monkey’s bid-marker up the budget-bar, to place higher bids, and backwards movements of the joystick moved the bid-marker down. The bid-marker could not move beyond the upper and lower limits of the budget-bar. Both monkeys had been trained to move the bid-marker in a separate ‘target-task’ (Ch. 4.1).

The computer-bid was indicated by a similar green bid-marker which only appeared once the monkey had placed his bid. If the computer’s bid was lower than the monkey’s bid, then the area of the budget-bar below the computer bid-marker was occluded. Computer bids were randomly generated from a Beta (4,4) distribution on each trial. This distribution of computer-bids was used as the monkeys’ values for the juice-rewards were closer to the centre of the bidding range, and, compared to a uniform distribution of computer bids, the Beta (4,4) distribution increases the costs of deviating from optimality for all rewards over most of the bidding range (fig. 2.6).

Scale-lines were presented on the budget-bar to help the monkeys better identify the relative positions of the bid-markers and the relative distances between them to further aid the learning of task contingencies.

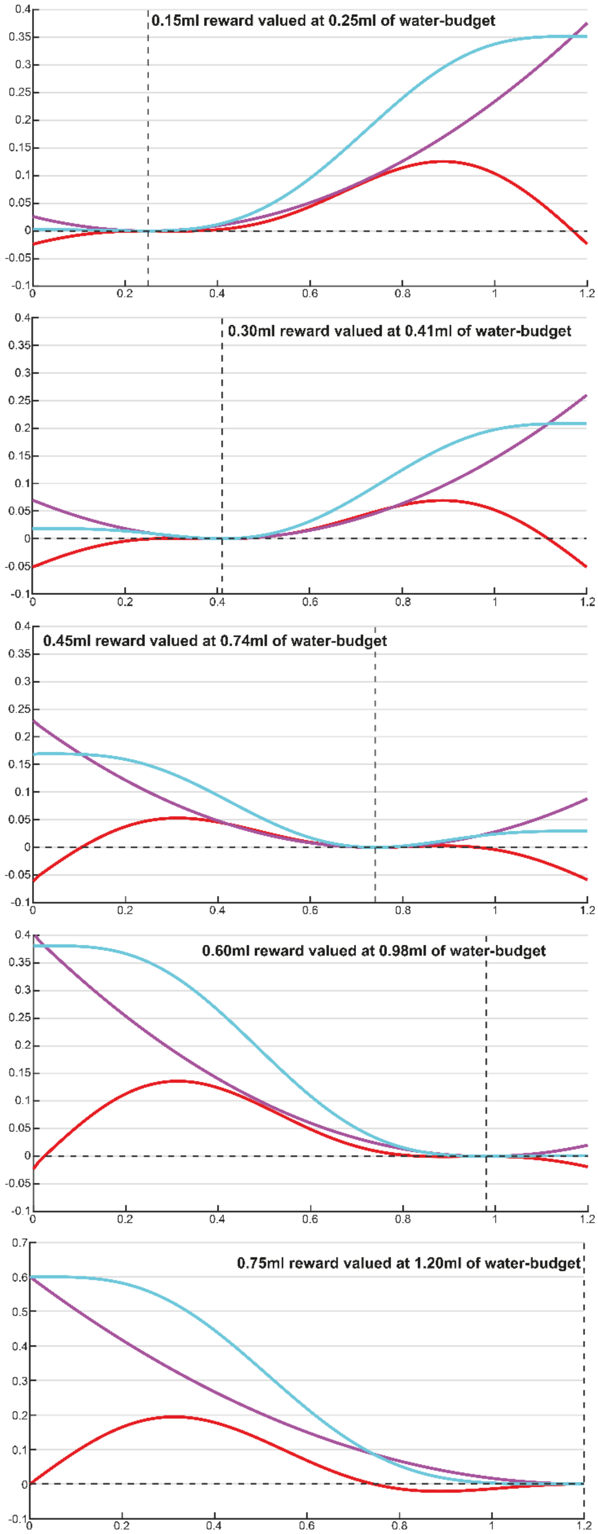


Fig. 2.6) The ECM is given for each juice-reward under both uniform (purple) and Beta (4,4) (blue) computer-bid distributions. The Y-axis gives the ECM and the X-axis gives the bid, both in millilitres of water budget. The red line shows the ECM difference given by subtracting the uniform ECM from the Beta (4,4) ECM. The Beta (4,4) ECM is larger for most of the bidding range for all rewards. Dotted vertical lines show the optimal bid and the values shown here for the different juice-rewards were taken from the BCb data of Ulysses (Ch. 3.2).

On each trial, the magnitude of the monkey's bid was calculated by finding the pixel position of the centre of their bid-marker, $\boldsymbol{Mp}$, and calculating its proportional height relative to the budget-bar height. Thus, where the pixel position of the top of the budget-bar is $\boldsymbol{Tp}$ and the pixel position of the bottom of the budget-bar is $\boldsymbol{Bp}$, the value of the bid is given by:

$$\boldsymbol{Bid} = \frac{(\boldsymbol{Mp} - \boldsymbol{Bp})}{(\boldsymbol{Tp} - \boldsymbol{Bp})} \times \mathbf{1.2}\boldsymbol{ml}$$

The computer's bid-marker position was found by starting with a known computer-bid that had been randomly generated for each trial and using the above relationship to find the appropriate central position for that bid.

<u>Trial progression:</u>

The monkey sat in a primate chair with an attached joystick and faced a screen on which task stimuli appeared. A spout positioned at the level of the monkey's mouth delivered both water-budget and juice-rewards.

The start of each trial was signalled by the appearance of a white cross in the centre of the screen (fig. 2.7). Following this, at the start of the 'Offer' epoch, a randomly selected reward-fractal was presented alongside the full budget-bar, and the starting position of the monkey's bid-marker was displayed using a dark red bid-marker. Figure 2.6 shows a version of the BDM task using a bottom starting position of the bid-marker (BS-BDM), but monkeys were also trained in versions of the task where the marker started at the top of the budget-bar (TS-BDM), or, at a completely random position within the budget-bar (RS-BDM). During this 'Offer' epoch the monkey had to maintain hold of the joystick and keep it centred. Any release of the joystick or movement outside of a central tolerance window (2% of the maximum deflection magnitude of the joystick in each direction)[*] would lead to 'no-hold' or 'not-centred' errors respectively (all errors led to display of a blue time-out error screen for the remaining duration of the trial plus three seconds, this was followed by repetition of the trial. Error trials were unrewarded).

[*] See Appendix 2 for a detailed description of joystick controls.

After a variable exponentially distributed delay (or 'jitter') with a flat hazard function[56] - to prevent the monkey from predicting the timing of the start of the 'Choice' epoch - the bid-marker changed colour to bright red, signalling that the joystick could now be moved. The monkey then had 6s to move the bid-marker to their desired position and made a bid by stabilising the bid-marker in this position for 250ms. If the monkey failed to stabilise the joystick within the 'Choice' epoch, then the trial ended with a 'no-choice' error.

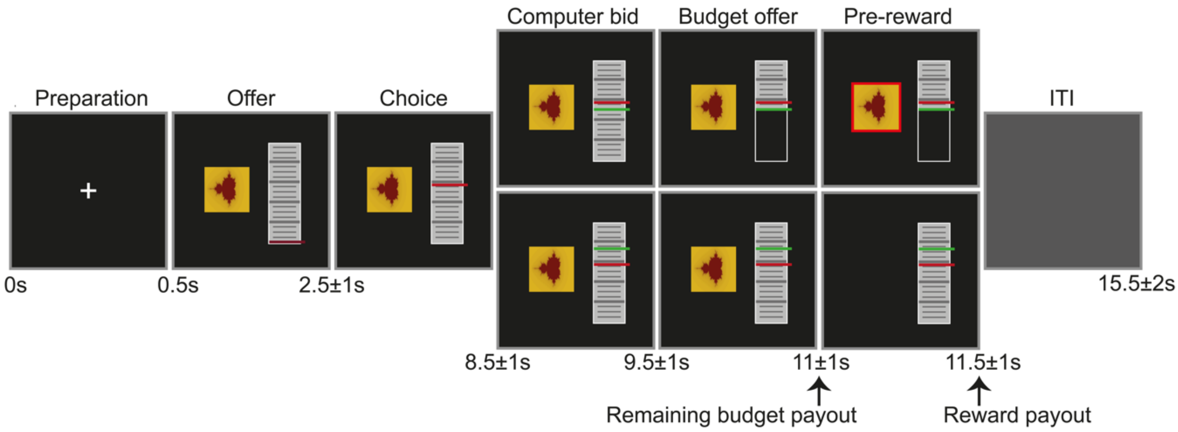


Fig. 2.7) Trials started with a 0.5s 'Preparation' epoch to orient the monkey to the screen. During the 'Offer' epoch both the reward-fractal and the budget-bar were shown for 2±1s (variable with a flat hazard function) alongside the starting position of the monkey's bid-marker (dark red). At the start of the 'Choice' epoch the bid-marker changed shade to bright red to signal that it could now be moved: the monkey then had 6s to move the bid-marker using the joystick. If the bid-marker was held in any given position for 250ms then it could no longer be moved, and the bid-marker position was taken as the monkey's bid for that trial. Regardless of when the joystick was stabilised, the duration of the 'Choice' epoch remained constant. Following this, the computer bid-marker appeared, and, depending on the relative position of the monkey's bid, the auction would either have been won (top panels) or lost (bottom panels). If the auction was lost then the monkey received all of the initial water-budget at the end of the 'Budget offer' epoch, but no juice-reward at the end of the 'Pre-reward' epoch. If the auction was won, then the paid portion of the water-budget was occluded at the beginning of the 'Budget offer' epoch and the remaining water-budget was received at the end of this epoch. A red border then appeared around the reward-fractal and the associated volume of juice-reward was received 0.5s later, at the end of the 'Pre-reward' epoch. Trials were separated by a variable inter-trial interval (ITI) of 4±1s, with random durations distributed by a truncated exponential function. The duration of each epoch as well as the overall trial duration were not affected by whether the auction was won or lost, and the timings of delivery of juice-reward and water-budget were held constant relative to the stimulus changes.

Following the ‘Choice’ epoch, the randomly generated computer-bid was presented, and, depending on its position relative to the monkey’s bid-marker, the auction was either won (fig. 2.7 – top panels) or lost (fig. 2.7 – bottom panels).

If the monkey won the auction, then the area below the computer-bid was occluded in the ‘Budget offer’ epoch, indicating both the paid and remaining volumes of water-budget to be delivered 2s later. After delivery of the remaining water-budget, the reward-fractal was surrounded by a red border and the associated volume of juice-reward was delivered 0.5s later.

On the other hand, if the monkey lost the auction, then budget-bar remained un-occluded and the full 1.2ml of water-budget was delivered at the same time as the monkey would have received the remaining water-budget in the case of a win. The reward-fractal then disappeared at the beginning of the ‘Pre-reward’ epoch to indicate that the auction was lost and that no juice-reward was to be delivered.

The delays between stimulus events and the delivery of remaining water-budget and juice-reward were held constant to maintain constant effects of *temporal discounting* on the values of the two rewards. Thus, the choice epoch length was kept constant regardless of the time at which the monkey’s bid-marker was stabilised and the relative timings of water-budget and juice-delivery were always the same.

The behavioural requirements relating to the control of the joystick (whose violation led to ‘no-hold’, ‘not-centred’, and ‘no-choice’ errors) were implemented to ensure that the monkey was paying attention to the task and was making purposeful bids. Similarly, the variable delay, or ‘jitter’ that was introduced into the ‘Offer’ epoch ensured that the monkey could not accurately predict the start of the choice epoch, and therefore had to keep attending to the on-screen events throughout this period. Without such a control monkeys could prepare and execute a movement at the correct time without attending to the screen[*].

[*] Monkeys were observed to follow such a ‘lazy’ strategy in previous versions of the task that did not implement such a ‘jitter’. It is possible that the low costs of deviating from optimality in the BDM (Ch. 2.1) could allow for such behaviour to develop unless attention is actively encouraged.

Experimental design:

For each session of testing, monkeys completed 200 correct trials of the BDM task that has been described above. Trials were tested in blocks of 50 correct trials each, with the monkeys taking a 5-minute break between blocks. The number of trials in a single session was held constant to control for the overall availability of the mango juice-reward and water-budget and to maintain a consistent reward rate.

Training sessions were conducted on separate days and on average both monkeys consumed 184.6ml (SD = 5.64)[*] of water-budget and 61.2ml (SD = 3.77) of juice-reward per day - a total fluid consumption of ~245.8ml for both monkeys. Where the monkey's total volume of liquid consumed did not meet our daily target of 30ml/kg (though the Home Office licence states a minimum of 20ml/kg), they were given a bolus of water at the end of the session to meet the target. This was rarely needed for Vicer, who at 7.9kg usually got ~31.1ml/kg. On the other hand, Ulysses, weighing 10.8kg, got ~22.8kml/kg and thus usually received a bolus of ~80ml of water at the end of each session.

Three different starting positions of the monkeys' bid-marker were tested, with 10 BDM sessions being tested for each of the bottom-start position (BS-BDM), top-start position (TS-BDM), and random-start position (RS-BDM) versions of the task. Different starting positions of the monkey's bid-marker were used to rule out possible confounds that could result in rank-ordered bidding by means that were independent of any response to the task structure or payoffs.

First, the TS-BDM was used to rule out the possibility that the ordering of bids in the BS-BDM was a result of increased effort expenditure when the good on offer had a higher value (and there was therefore more potential value on offer in the trial). This would lead to bids reflecting preferences by virtue of increased movement vigour[57,58], however, the top-start position reverses the relationship between effort expended and the magnitude of the bid such that greater backwards movements of the joystick would lead to lower bids (Ch. 5.5). Moreover, training the monkeys in the TS-BDM task meant that they could move the joystick in a well-controlled manner in both forward and backward directions, preparing them for the RS-BDM task.

---

[*] SD is used as an abbreviation for 'standard deviation' throughout.

Second, the RS-BDM was used to rule out not only a vigour effect, but also the possibility that monkeys were simply learning a specific distance of movement of the bid-marker for each of the top and bottom starting positions. Thus, the random start position ensured that monkeys were making intentional bids to a specific location on the budget-bar on each trial, and provided the most convincing evidence that monkeys had learnt the BDM task.

Initially, we had attempted to test the BCb task alongside the BDM in interleaved blocks of trials so that the values elicited by the two mechanisms could be compared within the same day. However, it was found that the monkeys' performance greatly improved when they did not have to switch between tasks (Ch. 5.6) – indeed, switching between task rules and different actions is difficult for both human and monkey subjects[59,60].

Therefore, BCb and BDM trials were separated in the latest version of the task presented here, with 5 BCb sessions preceding BDM testing and 5 sessions conducted immediately after. Values were inferred from the preceding, succeeding, and pooled (both taken together) sets of BCb sessions, so that any changes in the values for the juice-rewards over the course of BDM training could be accounted for. Nevertheless, juice-reward values were relatively stable.

The BCb task, which was used as an external check of the monkeys' performance in the BDM, is outlined in the following section, and the results of the performance of both monkeys in each of these tasks are presented in Chapter 3.

2.3 – Binary-choice-bundle (BCb) task design

The Binary-choice-bundle (BCb) task was designed to be as close an analogue to the BDM as possible. The rationale for using a bundle of juice-reward and a variable amount of water-budget and the full budget alone as the two reward options is explained in detail in Chapter 5, Section 3, where the development of the first version of this task is described. For now, suffice to say that the comparison of these two options is akin to the situation that the monkey faces in the BDM – their BDM bid should be equal to the maximum amount of budget that they would be willing to trade in order to receive the good: it is the point at which they are indifferent between the whole of their budget on one hand, and the good alongside the remaining budget on the other.

By varying the amount of water-budget offered in the bundle, the indifference point between the bundle and the budget-only option can be found. The value of this indifference point gives the minimum amount of remaining water-budget that the monkey must have alongside the juice-reward such that they are indifferent between that bundle and the full 1.2ml water-budget. Therefore, the equivalent BDM bid is inferred by subtracting the magnitude of water-budget in the bundle at the indifference point from this maximum water-budget volume of 1.2ml.

The reward indicating stimuli that were used in this BCb task were identical to those used in the BDM (fig. 2.8). One of the reward-fractals of the BDM task appeared alongside a budget-bar to indicate the bundle option. A green line (identical to the computer bid-marker of the BDM task) showed the amount of water-budget available in the bundle, and this varied from trial to trial. The monkeys had already been taught to expect that the area below this green line would be occluded at the beginning of the 'Budget offer' epoch, and this stimulus change was implemented at the same time as it would have been in the BDM task, so that the BCb task would be as similar to it as possible: both in terms of the timing of stimulus changes and their relationship to reward delivery. Similarly, the budget-only option was a budget-bar with the green line positioned at the very bottom of the bar, indicating that no water-budget would be subtracted from this option. Bundle and budget-only options were presented randomly on the left or right sides of the screen.

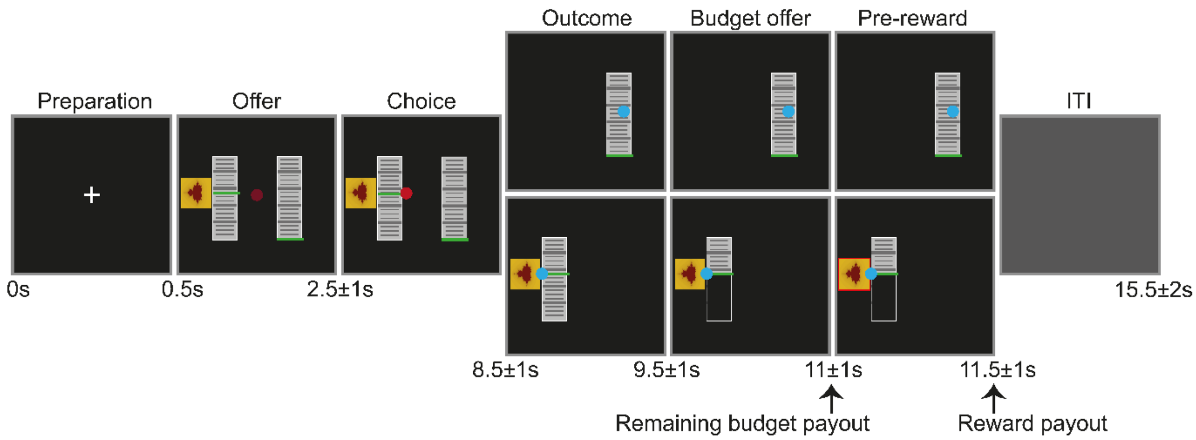


Fig. 2.8) Trials started with a 0.5s ‘Preparation’ epoch to orient the monkey to the screen. During the ‘Offer’ epoch, both the bundle and budget-only options were shown on random sides of the screen, alongside a dark-red choice-marker, for 2±1s (variable with a flat hazard function). A green line, (identical to the computer bid-marker of the BDM task) indicated the amount of water offered in the bundle. At the start of the ‘Choice’ epoch, the monkey’s choice-marker changed shade to bright red, signalling that it could now be moved: the monkey then had 6s to move the choice-marker to one of the equidistant options, and the choice-marker changed colour to light blue once it was in a position that would be recognised as constituting a valid choice, otherwise the trial ended with a ‘no-choice’ error. On correct trials, the monkey either chose the budget-only option (top panels) or the bundle option (bottom panels), with the unchosen option disappearing at the beginning of the ‘Outcome’ epoch, immediately after the ‘Choice’ epoch had ended. If the monkey chose the budget-only option, then they received all 1.2ml of water-budget at the end of the ‘Budget offer’ epoch, and no reward at the end of the ‘Pre-reward’ epoch. If the bundle was chosen, then the water-budget in the bundle would be indicated by the occlusion of the area below the green line at the start of the ‘Budget offer’ epoch, with the water-budget in the bundle being delivered at the end of this epoch. Finally, the reward-fractal in the bundle was surrounded by a red-border and receipt of the associated volume of mango juice-reward occurred 0.5s after. Trials were separated by a variable inter-trial interval (ITI) of 4±1s, with random durations distributed by a truncated exponential function.

However, instead of a bid-marker, the monkey controlled a red ‘choice-marker’ using left/right movements of the joystick. As in the BDM, the monkey had to maintain hold of the joystick throughout the ‘Offer’ epoch or the trial would end with a ‘no-hold’ error. During this epoch, the monkey also could not move the joystick outside of a central tolerance window (2% of the maximum deflection magnitude in each direction), else the trial would end with a ‘not-centred’ error. Once the choice-marker had changed from a dark red to a bright red colour, the monkey could move the joystick and choice-marker and the start of this ‘Choice’ epoch also followed an exponentially distributed variable delay with a flat hazard function just as in the BDM.

Once the ‘Choice’ epoch had started, the monkey could make left and right movements of the joystick to move the choice-marker over their preferred option, and this resulted in a colour change of the marker to blue, indicating that the monkey had selected a valid option. If the monkey failed to do this within the 6s ‘Choice’ epoch, then the trial ended with a ‘no-choice’ error. Just as in the BDM, all errors ended the trial and led to a time-out screen of duration equal to the remaining trial duration plus 3 seconds.

The unchosen option disappeared at the end of the ‘Choice’ epoch, and this was soon followed by occlusion of the ‘paid’ portion of the budget-bar at the end of the ‘Outcome’ epoch. The monkey’s chosen option was delivered with timings identical to those of the BDM task to control for any effects of temporal discounting. Thus, whatever water-budget was on offer in the chosen option was always delivered 2.5s after the ‘Choice’ epoch had ended, and if the chosen option contained juice-reward, then the reward-fractal was surrounded by a red border just as in the BDM task and the associated volume of mango juice was delivered 0.5s later, at the end of the ‘Pre-reward’ epoch.

Trials were presented in blocks of 10 trials for each juice-reward volume such that each volume of water-budget in the bundle was tested between 0 and 1.2ml in steps of 0.133ml. The order in which these water-budget offers in the bundle option were presented was randomised.

Just as in the BDM task, sessions were composed of 200 correct trials, with monkeys taking a 5-minute break every 50 correct trials - again, maintaining similar rates of reward consumption between the BDM and BCb tasks. Overall, this meant that there were 40 trials for each of the 5 juice-reward volumes, with 4 trials at each water-budget volume in the bundle option for each of these juice-rewards.

3

BDM results

Two monkeys, Ulyyses and Vicer, were tested in 3 different conditions of the BDM task. These conditions were identical except for the starting position of the monkey's bid-marker: a bottom starting position condition (BS-BDM) was tested first, followed by a top starting position task (TS-BDM), and finally a random starting position condition (RS-BDM). Different starting positions were tested to control for any potential motor effects, and to decorrelate the effort expended and the value of the bid. The key version of the task is the RS-BDM, which provides the best control of movement in relation to the placement of bids, and was tested when the monkeys were at their most experienced in the BDM. Nevertheless, the results of all three conditions have been presented for most analyses in this chapter, and they represent the performance of the monkeys on the latest, most refined version of the BDM task - tested between 01/06/17 and 01/09/17 (see fig. 5.33 for a summary of how these sessions stand in relation to previous training tasks).

Each of these BDM conditions was composed of 10 sessions of 200 correctly completed trials each. On each trial one of 5 mango juice-rewards was randomly presented (either 0.15ml, 0.30ml, 0.45ml, 0.60ml or 0.75ml) and the monkey had to place a bid in terms of water-budget for that reward. The results of bidding in both monkeys in all 3 of the conditions are presented in the following section, with an analysis of the consistency of the monkeys' bids and various factors that might influence their bidding. The monkeys both showed similarly meaningful performance in all 3 task conditions, and bid according to their order of preference for the juice-rewards.

To assess the monkeys' bids in more detail they were each tested in 5 binary-choice-bundle (BCb) sessions preceding BDM testing (Pre-BDM) and 5 BCb sessions immediately after the RS-BDM task had been completed (Post-BDM). Thus, any change in the monkeys' values for the mango juice-rewards over the period of BDM testing could be observed and accounted for. More importantly, the subjective values elicited in the BCb task were used to assess the optimality of the monkeys' bids in the BDM (Ch. 3.2) - as bids could reflect the order of preference without faithfully reproducing the same subjective values as the BCb task.

Finally, the chapter concludes with a discussion of the results and the viability of the BDM mechanism for use in monkeys (Ch. 3.3).

3.1 - The BDM reveals monkeys' preferences reliably

In each of all 30 BDM testing sessions, monkeys' mean bids reflected their order of preference for the rewards: their mean bids were always higher for higher magnitudes of mango juice. The mean bids of both Ulysses and Vicer in each session of all three conditions are shown in figures 3.1 and 3.3 respectively. Boxplots are also shown for every session to give an impression of the distribution and consistency of bidding (figure 3.2 for Ulysses and 3.4 for Vicer).

Error trials have been excluded from the following analyses as they were only in place to ensure that the monkeys were attending to the task while placing bids - there was little of interest in error trials themselves. However, it is worth noting that while Ulysses' overall error rate was only 6.7% across all sessions, it was 31% for Vicer, though this was largely due to long strings of ignored trials. This may indicate a lack of motivation and engagement during the task which may have been reflected in Vicer's more inconsistent bidding (see Ch. 3.3 for a discussion of differences in the overall performance of the monkeys).

While there was some variability in the bids for a given reward across different sessions, the size of this effect was small when compared to the effect of the juice-reward - which should be the main factor influencing the monkeys' bids[*]. For both monkeys, 2-way ANOVAs with factors of 'session' (10 levels) and 'reward' (5 levels) for each condition did find significant effects of both factors, and their interaction, on the monkeys' bids, however, the size of this effect was far greater for the 'reward' factor than for 'session' or for the 'reward*session' interaction (Table 3.1 for Ulysses and Table 3.2 for Vicer).

For example, in the BS-BDM for Ulysses, the identity of the juice-reward explained 80.2% of the variance in the monkey's bids, while session effects[†] only explained 1.7% of the variance. The largest session effect was observed in the TS-BDM condition for Vicer, in this case the reward explained 61.4% of the variance in bids while session effects accounted for 5.1% - this effect can be observed qualitatively in

[*] In ANOVA analyses the rewards are taken as 5 different categories, and so the subjective values of juice-rewards and their volumes can be used interchangeably.
[†] Effect sizes were measured using eta squared or η^2 - calculation of η^2 is given in Appendix 3.

the pattern of Vicer's TS-BDM, where bids are more concentrated at the centre of the bidding range in the early sessions and become more differentiated over the course of testing (fig. 3.3 and fig. 3.4). This inconsistent performance in the TS-BDM for Vicer may have been due to the fact that he had considerably less experience than Ulysses in using a top-starting position of the bid-marker.

Importantly, mean bids for individual juice-rewards were found to be significantly different to one another in individual sessions. For Ulysses, this was the case in each of the 30 sessions that were tested by a Bonferroni-corrected multiple-comparisons t-tests of bids for different juice-rewards (Table 3.3). For Vicer, this was true of 21 of the 30 sessions, with correct ordering of mean bids in the other 9 sessions but excessive overlap of bids for some adjacent rewards (Table 3.4) - nevertheless, this was limited to ambiguity between two rewards in all but session 6 of the BS-BDM condition, for which the mean bid for the 0.30ml could not be distinguished from those of either the 0.15ml or the 0.45ml juice-rewards.

Pooling across sessions of the same condition, the mean bids for each juice-reward were all found to be significantly different to one another - all $p < 0.05$, Bonferroni-corrected for multiple comparisons - for both monkeys in each condition (fig. 3.5).

Spearman's rank correlations of the reward volumes[*] and the monkeys' bids provided another measure of the extent to which the magnitude of bids was driven by the juice-rewards. Spearman's rank correlation was used as it does not assume a linear relationship between the reward volume and the monkey's bid, but rather assess how well this relationship can be described by a monotonic function. More consistent and well differentiated bidding leads to a higher value for Spearman's Rho, while a greater extent of overlap in the distribution of bids for different juice-rewards will lead to a lower value of Spearman's Rho.

Spearman's rank correlations of the monkey's bid and the juice-reward volume were significant on every individual session for both monkeys. The values of Spearman's Rho ranged from 0.87 to 0.94 (Mean = 0.91, SD = 0.017) in Ulysses (Table 3.3) and from 0.68 to 0.89 (Mean = 0.81, SD = 0.053) in Vicer (Table 3.4).

[*] Again, juice-reward volume or value can be used interchangeably as they will be given the same rankings in the analysis.

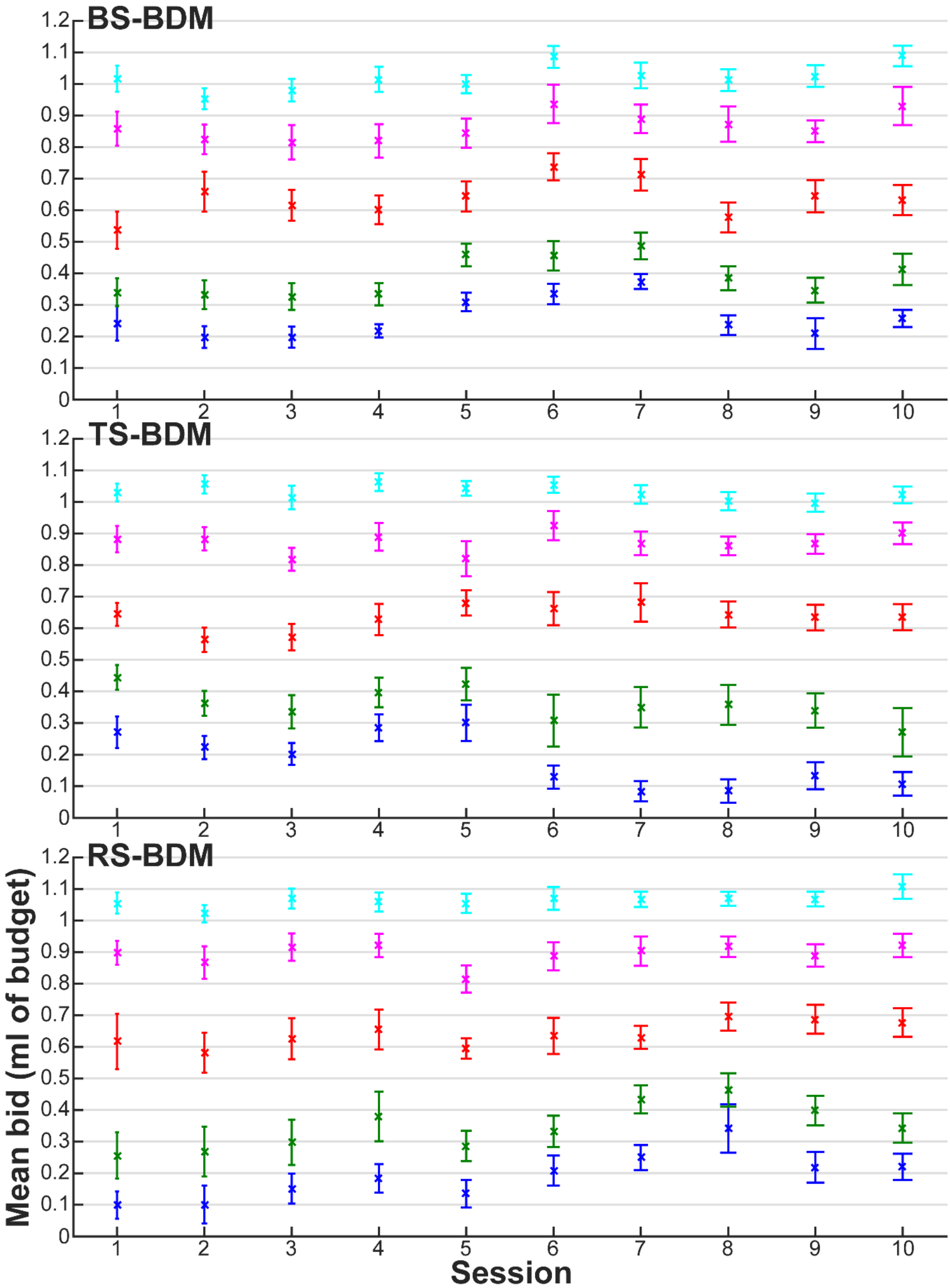


Fig. 3.1) The mean bids of Ulysses in each session of the bottom-start (BS-BDM), top-start (TS-BDM), and random-start (RS-BDM) versions of the task. Mean bids are shown for each of the 5 juice-reward volumes (0.15ml, blue; 0.30ml, green; 0.45ml, red; 0.60ml, magenta; 0.75ml, cyan). Error-bars are 95% confidence intervals of the mean.

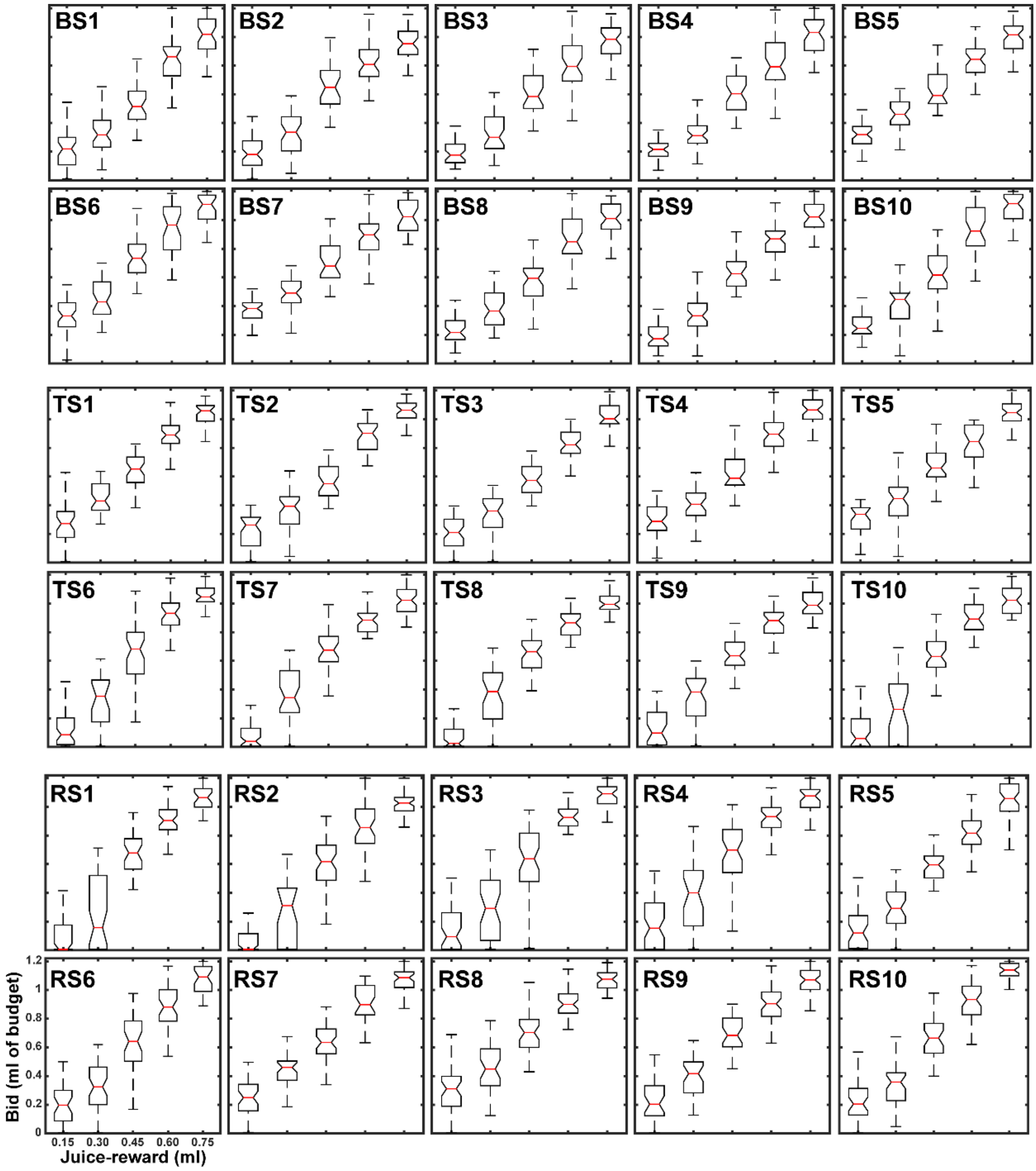


Fig. 3.2) As in figure 3.1, showing the bids of Ulysses for each reward in each session of the three conditions: bottom-start (BS), top-start (TS) and random-start (RS). Notches show 95% confidence intervals for the median, which is indicated by the red line in each boxplot. Boxes show the inter-quartile range and whiskers cover the range of bids. Bids were well differentiated in all 3 conditions.

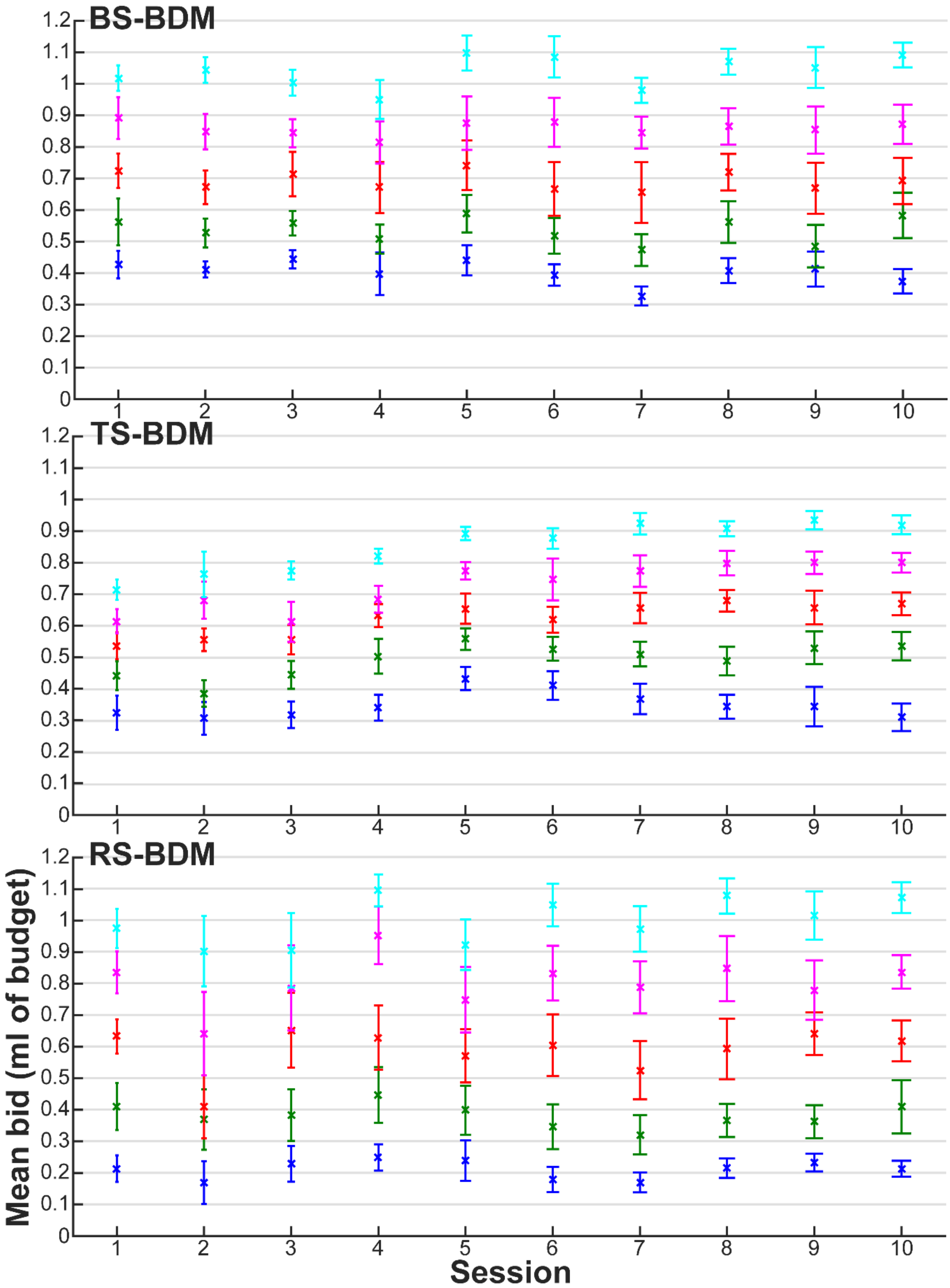


Fig. 3.3) The mean bids of Vicer in each session of the bottom-start (BS-BDM), top-start (TS-BDM), and random-start (RS-BDM) versions of the task. Mean bids are shown for each of the 5 juice-reward volumes (0.15ml, blue; 0.30ml, green; 0.45ml, red; 0.60ml, magenta; 0.75ml, cyan). Error-bars are 95% confidence intervals of the mean.

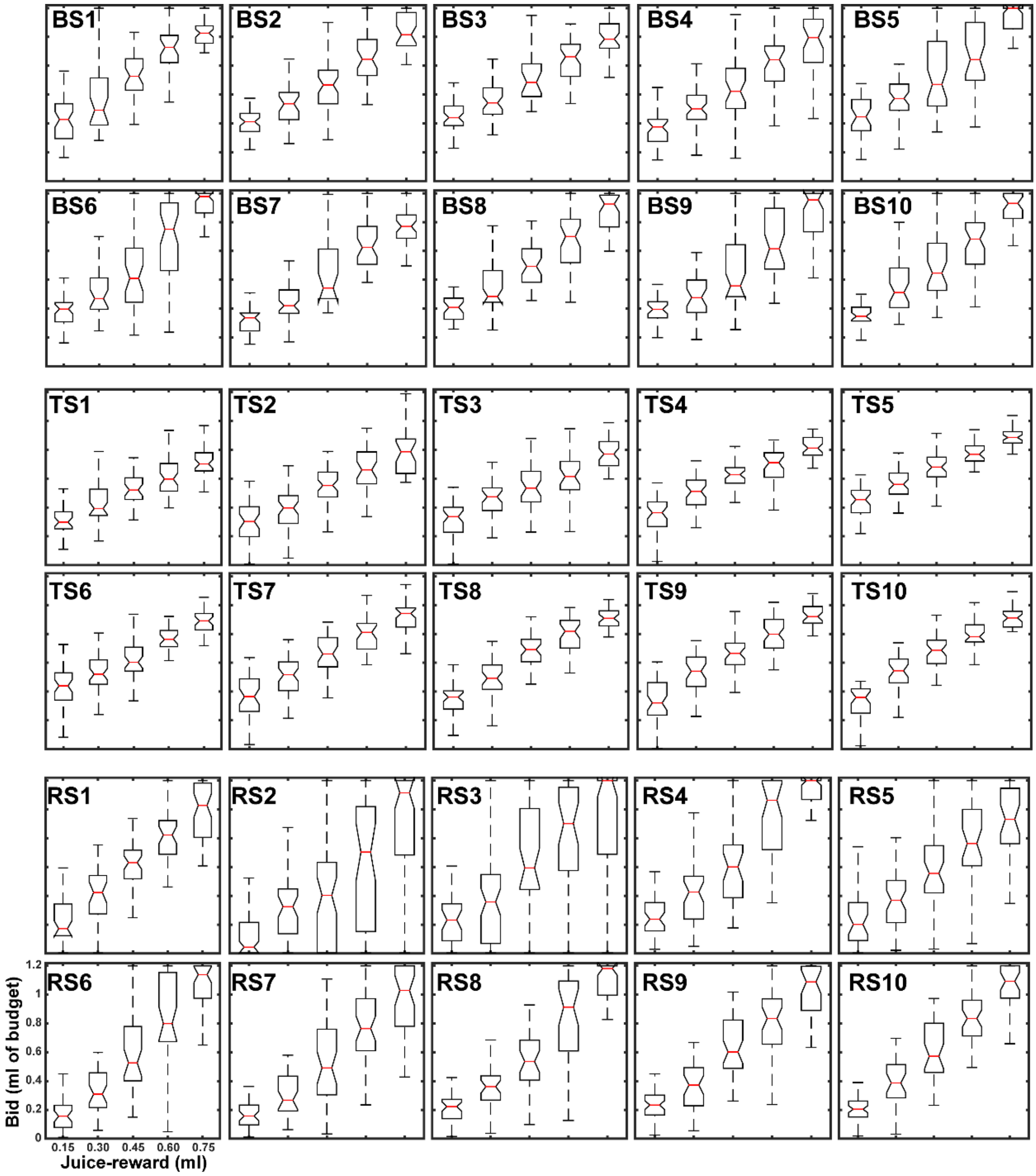


Fig. 3.4) As in figure 3.3, showing the bids of Vicer for each reward in each session of the three conditions: bottom-start (BS), top-start (TS) and random-start (RS). Notches show 95% confidence intervals for the median, which is indicated by the red line in each boxplot. Boxes show the inter-quartile range and whiskers cover the range of bids. Bids were well differentiated, but with greater variability in the random-start condition. Bids also covered less of the bidding-range in the early sessions of the top-start condition.

Table 3.1) 2-way ANOVA results of the effects of 'Session', 'Reward', and their interaction in Ulysses for each BDM condition: shown are the sum of squares (SS), degrees of freedom (d.f.), Mean Square (MS), F-statistic (F), p-value (p) and eta squared (η^2).

Condition	Factor	SS	d.f.	MS	F	p	η^2
BS-BDM	Session	3.53	9	0.39	22.25	2.63×10^{-36}	0.02
	Reward	163.51	4	40.88	2316.92	0	0.80
	Session*Reward	1.60	36	0.04	2.52	1.91×10^{-6}	0.01
	Error	34.40	1950	0.02			0.17
	Total	203.84	1999				
TS-BDM	Session	1.40	9	0.16	9.07	1.51×10^{-13}	0.01
	Reward	196.92	4	49.23	2866.24	0	0.82
	Session*Reward	3.07	36	0.09	4.97	8.92×10^{-20}	0.01
	Error	33.49	1950	0.02			0.14
	Total	241	1999				
RS-BDM	Session	2.95	9	0.33	14.87	1.37×10^{-23}	0.01
	Reward	206.76	4	51.7	2348.51	0	0.80
	Session*Reward	1.61	36	0.05	2.03	3.08×10^{-4}	0.01
	Error	42.92	1950	0.02			0.17
	Total	258.87	1999				

Table 3.2) 2-way ANOVA results of the effects of ‘Session’, ‘Reward’, and their interaction in Vicer for each BDM condition: shown are the sum of squares (SS), degrees of freedom (d.f.), Mean Square (MS), F-statistic (F), p-value (p) and eta squared (η^2).

Condition	Factor	SS	d.f.	MS	F	p	η^2
BS-BDM	**Session**	1.56	9	0.17	5.24	4.52×10^{-7}	0.01
	Reward	101.05	4	25.26	762.65	0	0.60
	Session*Reward	0.85	36	0.024	0.71	9.01×10^{-1}	0.01
	Error	64.59	1950	0.03			0.38
	Total	168.68	1999				
TS-BDM	**Session**	5.14	9	0.57	35.77	5.45×10^{-59}	0.05
	Reward	61.96	4	15.49	970.39	0	0.61
	Session*Reward	1.54	36	0.04	2.68	3.26×10^{-7}	0.02
	Error	31.13	1950	0.02			0.31
	Total	100.84	1999				
RS-BDM	**Session**	2.87	9	0.32	6.25	9.64×10^{-9}	0.01
	Reward	178.07	4	44.52	872.22	0	0.62
	Session*Reward	1.83	36	0.05	1	4.73×10^{-1}	0.01
	Error	99.53	1950	0.05			0.35
	Total	285.45	1999				

Table 3.3) For Ulysses, both the results of multiple-comparisons (MC) tests of the mean bids of all juice-rewards, and, the results of a Spearman's rank correlation between bids and juice-reward volumes, are given for each session. The pairwise comparisons for which the MC test was unable to reject the hypothesis that mean bids were the same for any two juice-rewards at an alpha level of 0.05 are given are given, otherwise, if all mean bids were different, the cell is left grey.

Condition	Session	MC results	Rho	P
BS-BDM	1		0.87	1.55×10^{-62}
	2		0.89	4.20×10^{-70}
	3		0.90	9.08×10^{-73}
	4		0.91	1.65×10^{-78}
	5		0.92	1.61×10^{-81}
	6		0.90	1.51×10^{-71}
	7		0.89	4.18×10^{-69}
	8		0.90	2.80×10^{-74}
	9		0.93	7.81×10^{-87}
	10		0.90	2.84×10^{-75}
TS-BDM	1		0.92	5.16×10^{-81}
	2		0.93	3.34×10^{-86}
	3		0.92	3.78×10^{-84}
	4		0.91	3.37×10^{-78}
	5		0.90	6.97×10^{-75}
	6		0.92	3.85×10^{-81}
	7		0.93	3.37×10^{-86}
	8		0.93	5.36×10^{-89}
	9		0.93	2.41×10^{-87}
	10		0.92	3.59×10^{-84}
RS-BDM	1		0.88	1.47×10^{-67}
	2		0.90	9.33×10^{-75}
	3		0.90	5.33×10^{-72}
	4		0.89	2.92×10^{-71}
	5		0.94	6.13×10^{-94}
	6		0.90	6.71×10^{-72}
	7		0.93	2.06×10^{-87}
	8		0.91	4.07×10^{-79}
	9		0.93	2.34×10^{-87}
	10		0.92	1.18×10^{-82}

Table 3.4) For Vicer, both the results of multiple-comparisons (MC) tests of the mean bids of all juice-rewards, and, the results of a Spearman's rank correlation between bids and juice-reward volumes, are given for each session. The pairwise comparisons for which the MC test was unable to reject the hypothesis that mean bids were the same for any two juice-rewards at an alpha level of 0.05 are given, otherwise, if all mean bids were different, the cell is left grey.

Condition	Session	MC results	Rho	P
BS-BDM	1		0.81	5.21×10^{-48}
	2		0.85	9.03×10^{-57}
	3		0.84	1.63×10^{-55}
	4		0.78	2.13×10^{-42}
	5		0.73	2.74×10^{-34}
	6	0.15ml/0.3ml & 0.3ml/0.45ml	0.75	9.98×10^{-37}
	7		0.82	6.99×10^{-51}
	8		0.80	4.52×10^{-46}
	9	0.15ml/0.6ml	0.73	6.74×10^{-35}
	10		0.78	1.82×10^{-41}
TS-BDM	1	0.45ml/0.6ml	0.73	1.21×10^{-34}
	2	0.6ml/0.75ml	0.79	7.20×10^{-44}
	3	0.45ml/0.6ml	0.79	2.09×10^{-44}
	4	0.45ml/0.6ml	0.83	4.88×10^{-52}
	5		0.88	1.41×10^{-64}
	6		0.81	7.00×10^{-47}
	7		0.84	2.70×10^{-55}
	8		0.88	5.13×10^{-67}
	9		0.85	2.90×10^{-57}
	10		0.89	3.65×10^{-69}
RS-BDM	1		0.86	3.38×10^{-59}
	2		0.75	9.83×10^{-38}
	3	0.6ml/0.75ml	0.75	8.52×10^{-38}
	4	0.6ml/0.75ml	0.80	2.57×10^{-45}
	5	0.15ml/0.3ml	0.68	3.66×10^{-28}
	6		0.81	4.61×10^{-48}
	7		0.81	2.16×10^{-47}
	8		0.84	4.39×10^{-55}
	9		0.86	7.99×10^{-59}
	10		0.88	6.40×10^{-66}

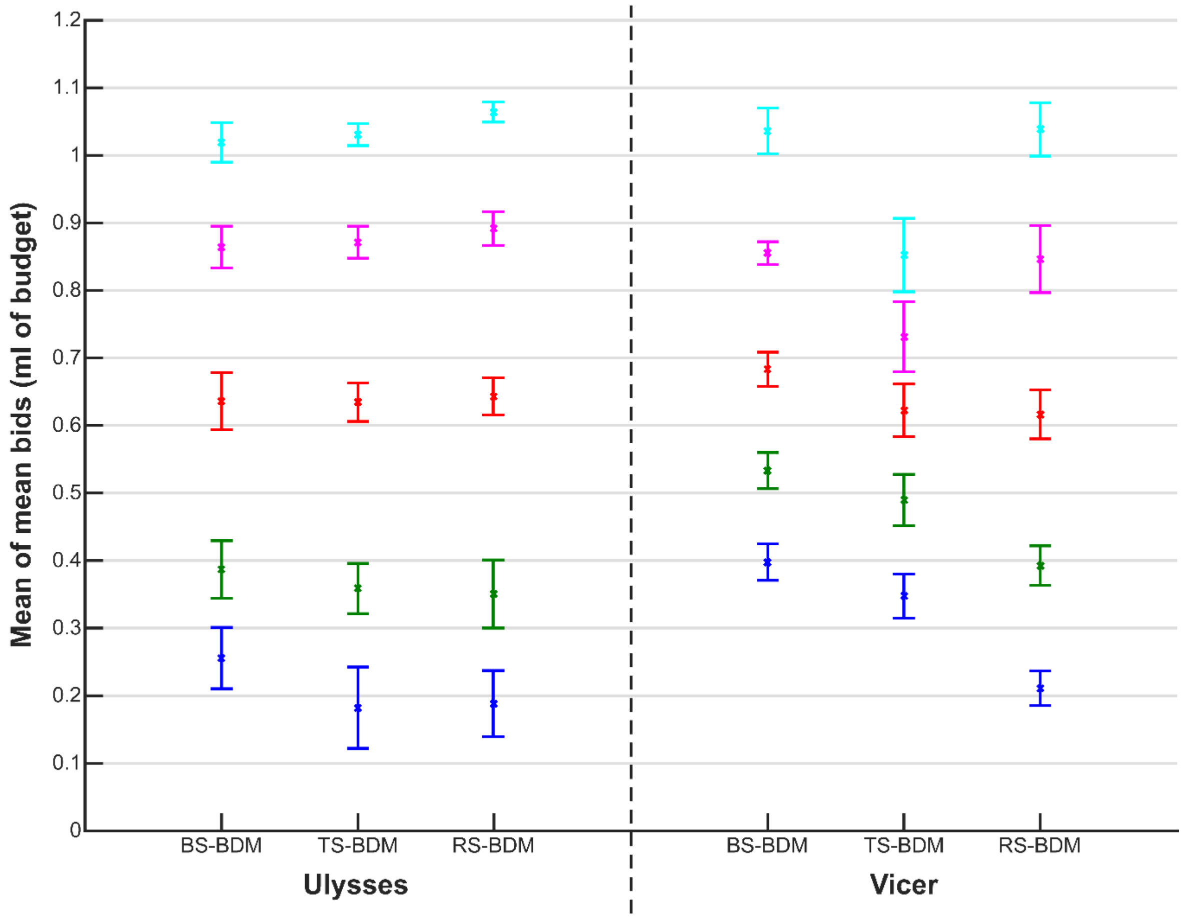


Fig. 3.5) The mean bids for each reward (0.15ml, blue; 0.30ml, green; 0.45ml, red; 0.60ml, magenta; 0.75ml, cyan) and in each bid-marker starting position, after pooling all sessions of the same condition for both monkeys. Ulysses' bids varied less between conditions, but he had more experience and may have been at/near an asymptotic level of learning.

After pooling all sessions of the same bid-marker starting position condition, 2-way ANOVAs of the effects of 'condition' and 'reward' found significant effects of both factors, and their interaction, on the monkey's bids for both monkeys. However, for Ulysses the effect size of the condition factor was vanishingly small, ~0%, with reward value and the error term accounting for 81.9% and 17.8% of the variance respectively, and the interaction of condition and reward accounting for only 0.3% of the variance. For Vicer, the effect of bid-marker starting condition was more pronounced, but still small compared to the effect of the reward: reward accounted for 58% of the variance, while condition and the reward-condition interaction accounted for 1.8% and 2.7% of the variance respectively - with 36.9% captured by the error term (Table 3.5).

Table 3.5) 2-way ANOVA results of the effects of bid-marker starting position 'Condition', 'Reward', and their interaction in both monkeys: shown are the sum of squares (SS), degrees of freedom (d.f.), Mean Square (MS), F-statistic (F), p-value (p) and eta squared (η^2).

Monkey	Factor	SS	d.f.	MS	F	p	η^2
Ulysses	**Reward**	576.38	4	144.09	6889.46	0	0.82
	Condition	0.30	2	0.15	7.18	7.71×10^{-4}	0.0004
	Reward* Condition	2.27	8	0.28	13.55	1.24×10^{-19}	0.003
	Error	125.18	5985	0.02			0.178
	Total	703.84	5999				
Vicer	**Reward**	329.01	4	82.25	2353.17	0	0.58
	Condition	10.41	2	5.21	148.94	7.49×10^{-64}	0.018
	Reward* Condition	15.62	8	1.95	55.86	3.94×10^{-88}	0.028
	Error	209.2	5985	0.03			0.369
	Total	566.41	5999				

Thus, both monkeys could reliably differentiate the 5 different juice-rewards, according to their order of preference, by bidding in individual sessions of the BDM task. Moreover, they could do this regardless of the starting position of their bid-marker, and even when that position was randomised on each trial. It is therefore difficult to account for this behaviour by means of the simple learning of a sensory-motor association - rather, both monkeys seemed to make specific and purposeful bids for different volumes of the mango juice-reward.

Moreover, the monkeys did not need to complete all 200 trials in a session before the different juice-rewards could be differentiated by their bids. For each session, we identified the first trial, T_{MC}, at which the mean bids were in order of preference, and, for which a Bonferroni-corrected multiple-comparisons t-test found a significant difference between the mean bids of all juice-rewards for that trial and for the 9 correct trials that followed it. This latter requirement ensured the reliability of the differentiation of bids at the first trial that was identified in this way. If 10 consecutive trials for which this held were not found, but the bids could be statistically differentiated by the end of the session, then the value of T_{MC} was taken as 200.

The mean value of T_{MC} did not differ significantly between the different conditions and was 113.4 (SD = 45.1) in the 30 sessions for which the mean bids for all rewards were found to be significantly different to one another for Ulysses,
and 179.5 (SD = 23.9) in the 21 sessions for which this was the case for Vicer. The mean bid for each reward at each correct trial, and the trial at which all bids could first be reliably distinguished by a multiple-comparisons test are shown in figure 3.6, overleaf.

These results suggest that the BDM can be used in monkeys to establish their preferences and potentially in fewer trials than would be possible using a binary-choice task. Notably, for Ulysses - who had more experience in the BDM task - the BDM could be used to establish the monkeys order of preference reliably in a single session, and, this could be done within ~113 trials on average - consider that in the BCb task in 100 trials each pairwise choice for each of the 5 juice-rewards would only be sampled twice.

Moreover, if fewer juice-rewards are tested then the differences in value between them can be kept larger and this leads to a far lower value of T_{MC}. For example, for Ulysses in a previous version of the BDM task using only 3 juice-rewards ([1.2-RS-BDM-B], Ch. 5.6) T_{MC} is found to have a mean value of 40.8 (SD = 13.7). Vicer performed similarly in a previous version which also only utilised 3 juice-rewards ([1.2-BDM-B], Ch. 5.6), with a mean T_{MC} of only 41.9 (SD = 17.4).

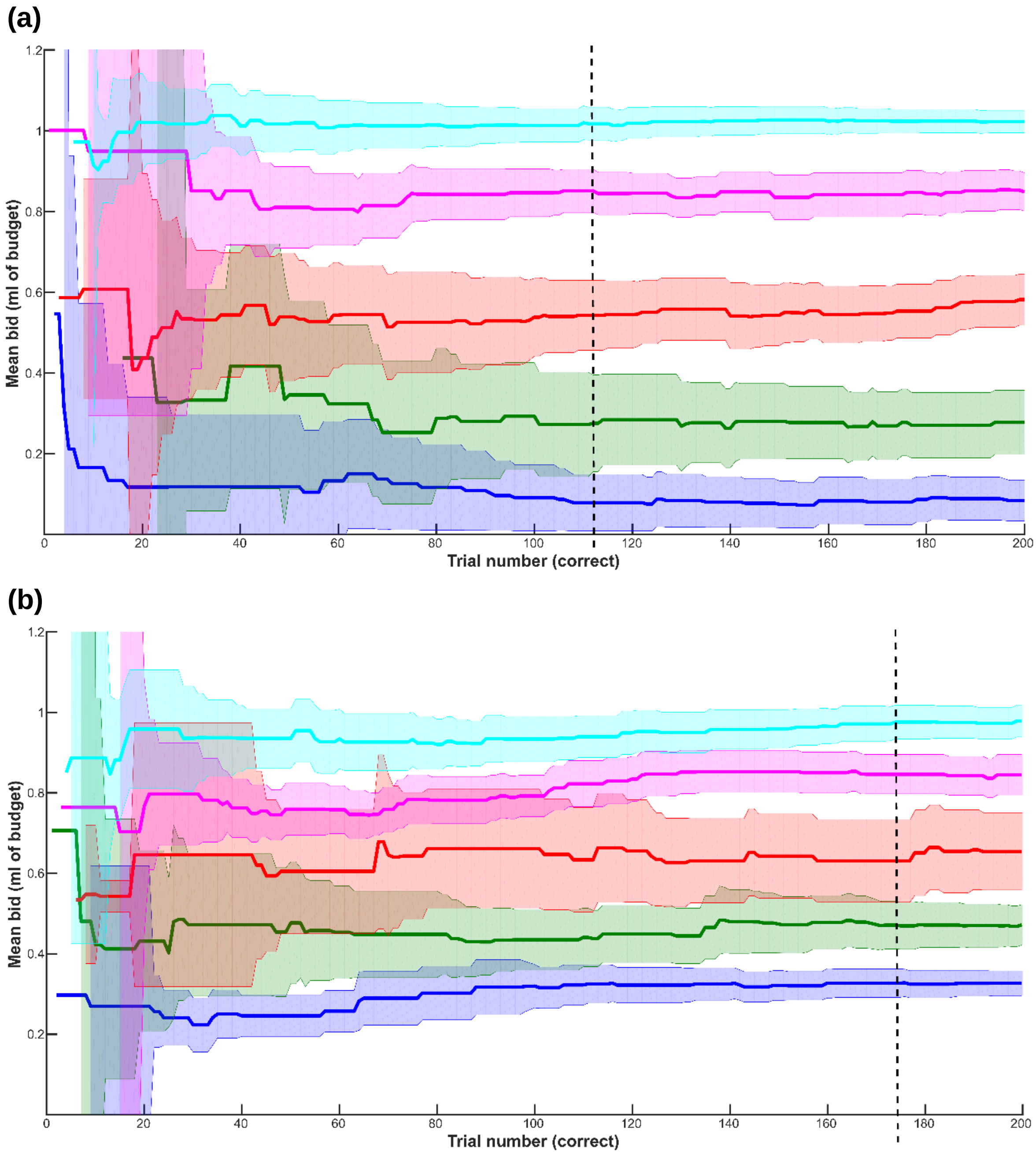


Fig. 3.6a) In session 2 of the RS-BDM, Ulysses' bids could be used to statistically differentiate all 5 juice-rewards (0.15ml, blue; 0.30ml, green; 0.45ml, red; 0.60ml, magenta; 0.75ml, cyan) in 113 correctly completed trials. Solid lines indicated the mean bid for each reward, shaded areas show the 95% confidence intervals of the means and the dashed vertical line shows the trial at which a Bonferroni-corrected multiple-comparisons t-test would find a significant difference in the mean bids for all juice-rewards up to that trial and the 9 trials that followed. **b)** On average, Vicer had to make more bids for each reward before they could be reliably differentiated. In session 7 of the BS-BDM this was possible only after 175 trials had been completed.

Finally, we wanted to examine various potential influences on the monkeys' bids and hypothesised that they may be modulated not only by the value of the juice-reward, but also by the value of reward that had been consumed up to, but not including, that trial, as well as the starting position of the bid-marker. We therefore carried out the following linear regression for after pooling all the BDM data for each monkey and representing all variables in terms of their z-scores:

$$\boldsymbol{Monkey\ Bid = B_0 + B_{Rval} * RewardValue + B_{ConVal} * ValueConsumed}$$
$$\boldsymbol{+ B_{Start} * BidMarkerStartPosition}$$

The results of this regression analysis for Ulysses showed a significant contribution of the reward value and the bid-marker start position, but no significant effect of consumption:

$\mathbf{B_0}$ = -1.55x10^{-13} (p = 1), $\mathbf{B_{Rval}}$ = 0.91 (p = 0),
$\mathbf{B_{ConVal}}$ = 0.0043 (p = 0.4), and $\mathbf{B_{Start}}$ = -0.033 (p = 2.37x10^{-9})

The value of $\mathbf{r^2}$ for this regression was 0.8195, however, conducting another linear regression with only reward value as a regressor, such that:

$$\boldsymbol{Monkey\ Bid = B_0 + B_{Rval} * RewardValue}$$

Gave an $\mathbf{r^2}$ of 0.8184, with $\mathbf{B_0}$ = -1.55x10^{-13} (p = 1) and $\mathbf{B_{Rval}}$ = 0.9 (p = 0).

These results suggested that for Ulysses, any effects of satiety or the bid-marker position were negligible, and that bids were primarily driven by the values of the rewards for which they were made.

The results of the first regression in Vicer showed a similarly primary effect of the reward value, but the influence of consumption was significant, and with a greater influence of the bid-marker starting position:

$$\mathbf{B_0} = 9.92\text{x}10^{-15} \text{ (p = 1)}, \mathbf{B_{Rval}} = 0.75 \text{ (p = 0)},$$
$$\mathbf{B_{ConVal}} = 0.017 \text{ (p = 0.042), and } \mathbf{B_{Start}} = -0.11 \text{ (p = } 1.93\text{x}10^{-38}\text{)}$$

The value of $\mathbf{r^2}$ for this regression was 0.58, and the second, simpler regression of only the monkey's bids and the reward values gave $\mathbf{r^2} = 0.57$, with coefficients:

$$\mathbf{B_0} = 1.03\text{x}10^{-14} \text{ (p = 1) and } \mathbf{B_{Rval}} = 0.76 \text{ (p = 0).}$$

Taken together, these results indicate that in both monkeys effects of factors other than the juice-reward's value were negligible and did little to explain the variance in the monkeys' bids beyond that which was explained by the juice-reward value alone. As such, satiety effects and the bid-marker starting position in the RS-BDM are not considered further, and the focus of further analyses was on the bids and their relation to the monkeys' values.

The results presented thus far are consistent with incentive compatible bidding, but they do not answer the question of whether the monkey's bids were equal to their subjective values for the rewards, or, how closely the bids reflected those values. To address this, the next section compares the values inferred in a supplementary BCb task with the monkeys' bids in the BDM.

3.2 - Comparing BDM and BCb values

Both monkeys were tested in 5 BCb sessions preceding BDM training (Pre-BDM) as well as in 5 BCb sessions immediately after the last RS-BDM session had been tested (Post-BDM). BCb testing was conducted before and after the BDM experiment so that any change in the monkeys' values for the juice-rewards could be detected and accounted for. In both monkeys, two-sample t-tests between the inferred optimal bids for each reward in the Pre-BDM and Post-BDM BCb tasks (fig. 3.7) found no significant differences in the mean inferred optimal bids (all $p > 0.05$).

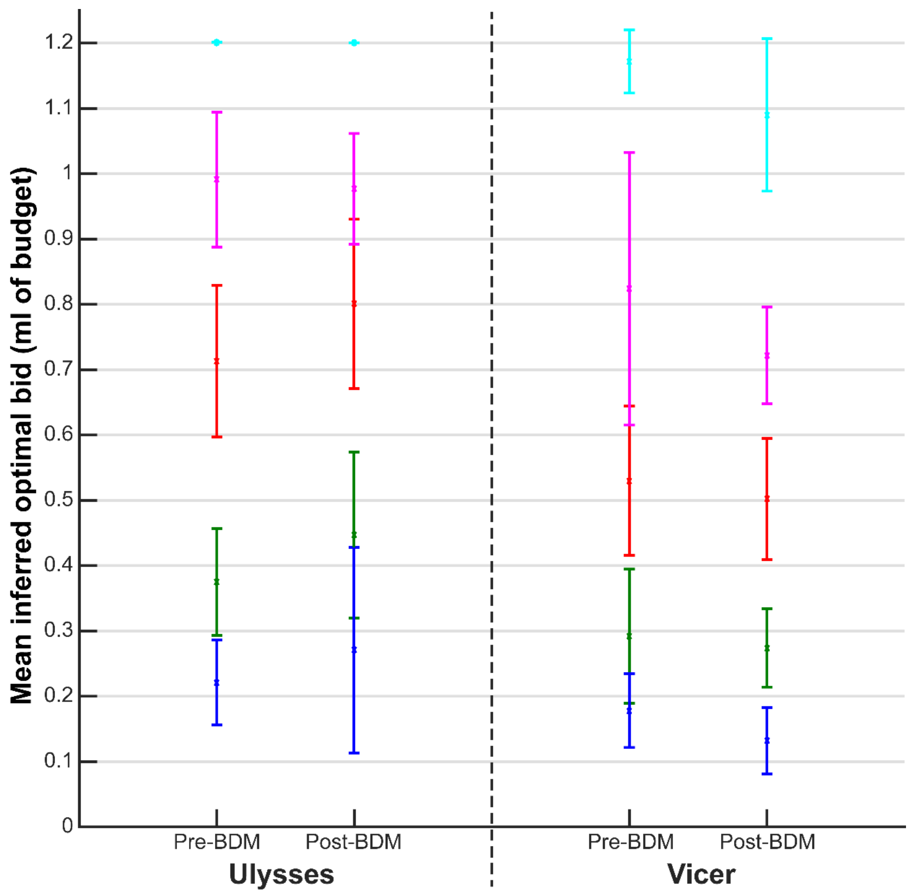


Fig. 3.7) The mean inferred optimal bid for each juice-reward (0.15ml, blue; 0.30ml, green; 0.45ml, red; 0.60ml, magenta; 0.75ml, cyan) from the 5 BCb sessions before and the 5 BCb sessions after BDM testing. Optimal bids were equal to the mean inferred values from the BCb task, except for the 0.75ml juice-reward in Ulysses, for whom the value was greater than the water-budget limit: in this case, the optimal bid is a maximum bid. Error-bars are 95% confidence intervals of the mean.

The inferred optimal bid was equal to the inferred value in every case except for the 0.75ml reward in Ulysses, for whom the value was greater than the water-budget limit. Therefore, for this reward, the optimal bid for Ulysses was a maximum bid.

This comparison of Pre-BDM and Post-BDM BCb sessions suggests that the optimal bids of both monkeys for each juice-reward should have been relatively stable across the BDM testing period. Moreover, it supports the pooling of Pre-BDM and Post-BDM BCb sessions (fig. 3.8). Because values did not vary significantly between the two blocks, they could be meaningfully pooled to attain more reliable estimates of the juice-reward values for use in the BDM analysis. There was no reason to use individual session data as these were collected on separate days to the BDM sessions anyway. Thus, individual BCb session values could not be meaningfully associated with specific BDM sessions, whereas pooling across BCb sessions could account for day-to-day variability in juice-reward values.

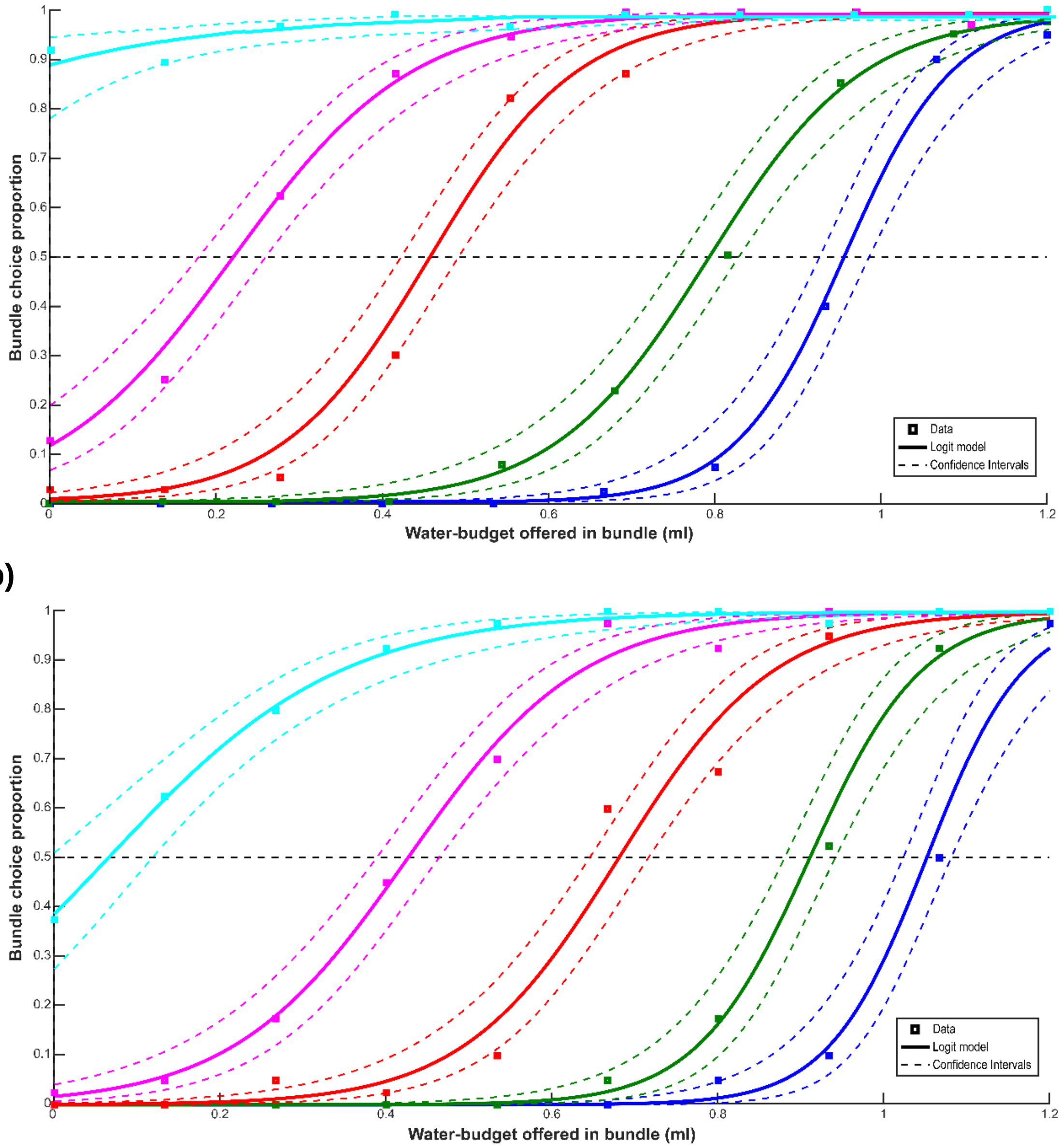
The indifference points and inferred optimal bids from the pooled BCb data are given in Table 3.6. The inferred juice-reward values were found by subtracting the amount of water-budget offered in the bundle at the point of indifference from the maximum water-budget[*V].

Table 3.6) The indifference points for the logistic regressions over the 10 pooled BCb sessions (fig. 3.8) are given for each monkey, as is the optimal bid (juice-reward value), which is equal to the maximum water-budget minus the value at the indifference point. Where the indifference point is less than 0ml, it is taken as 0ml for the purposes of inferring the optimal bid.

	Ulysses		**Vicer**	
Reward	**Indifference point**	**Inferred optimal bid**	**Indifference point**	**Inferred optimal bid**
0.15ml	0.95ml	0.25ml	1.05ml	0.15ml
0.30ml	0.79ml	0.41ml	0.91ml	0.29ml
0.45ml	0.46ml	0.74ml	0.68ml	0.52ml
0.60ml	0.22ml	0.98ml	0.43ml	0.77ml
0.75ml	<0ml	1.2ml	0.06ml	1.14ml

[*]Detailed logistic regression methods and the derivation of values from indifference points in a bundle-binary-choice task are given in Appendix 3.

(a)



(b)

Fig. 3.8a) Logistic regression for each of the 5 juice-rewards (0.15ml, blue; 0.30ml, green; 0.45ml, red; 0.60ml, magenta; 0.75ml, cyan) tested over all 10 BCb sessions in Ulysses. Indifference points for different juice-rewards are found where the coloured curves cross the black dotted horizontal line - see Table 3.6. For Ulysses, the value of the 0.75ml mango juice-reward did not fall within the 1.2ml budget-range. **b)** As for (a), but for the 10 BCb sessions of the second monkey, Vicer.

The monkey's optimal bid should correspond to the values inferred from the BCb task, as it is known to be incentive compatible and has been used to reliably estimate monkeys' subjective values in the past. Assuming that the BCb values were a valid estimate of the optimal bid, we could compare those values to the mean bids for the same reward in the BDM (fig. 3.9) and thereby assess the optimality of the monkey's bids.

We compared the mean bids for the juice-rewards from 10 sessions of the RS-BDM with their values from the 10 BCb sessions. For Ulysses, 2-sample t-tests showed significant differences between the BCb values and the mean bids for all except the 0.15ml and 0.30ml juice-rewards (both $p > 0.05$). Similarly, for Vicer, this analysis found a significant difference between the values and mean bids of all juices across the two tasks, except for the 0.6ml juice-reward ($p = 0.15$). The results of comparisons between individual BDM sessions and the pooled BCb values for each reward were similar (Tables 3.7 and 3.8) - note that no correction was applied for multiple comparisons, as this would make the tests less conservative in this case.

The differences between the mean bids in each condition and the optimal bids as inferred from the BCb values formed part of a general pattern of underbidding in Ulysses (fig. 3.10a) and overbidding in Vicer (fig. 3.10b).

However, while Ulysses' bids were relatively stable across the different conditions, Vicer showed an improvement as he progressed from the BS-BDM to the RS-BDM task and his mean bids for most of the juice-rewards approached their optimal value. Vicer's general improvement in bidding could also be seen in terms of the overall mean absolute distance of bids from the optimal bid, or absolute bid deviance (ABD) for all juice-rewards (fig. 3.11). The results of Bonferroni-corrected multiple-comparisons t-tests between the values of ABD in each condition showed a general trend of increasing proximity of bids to the optimal: the RS-BDM ABD was significantly less than that of the TS-BDM or BS-BDM, and the difference between the TS-BDM and BS-BDM ABD was also close to statistical significance ($p = 0.056$).

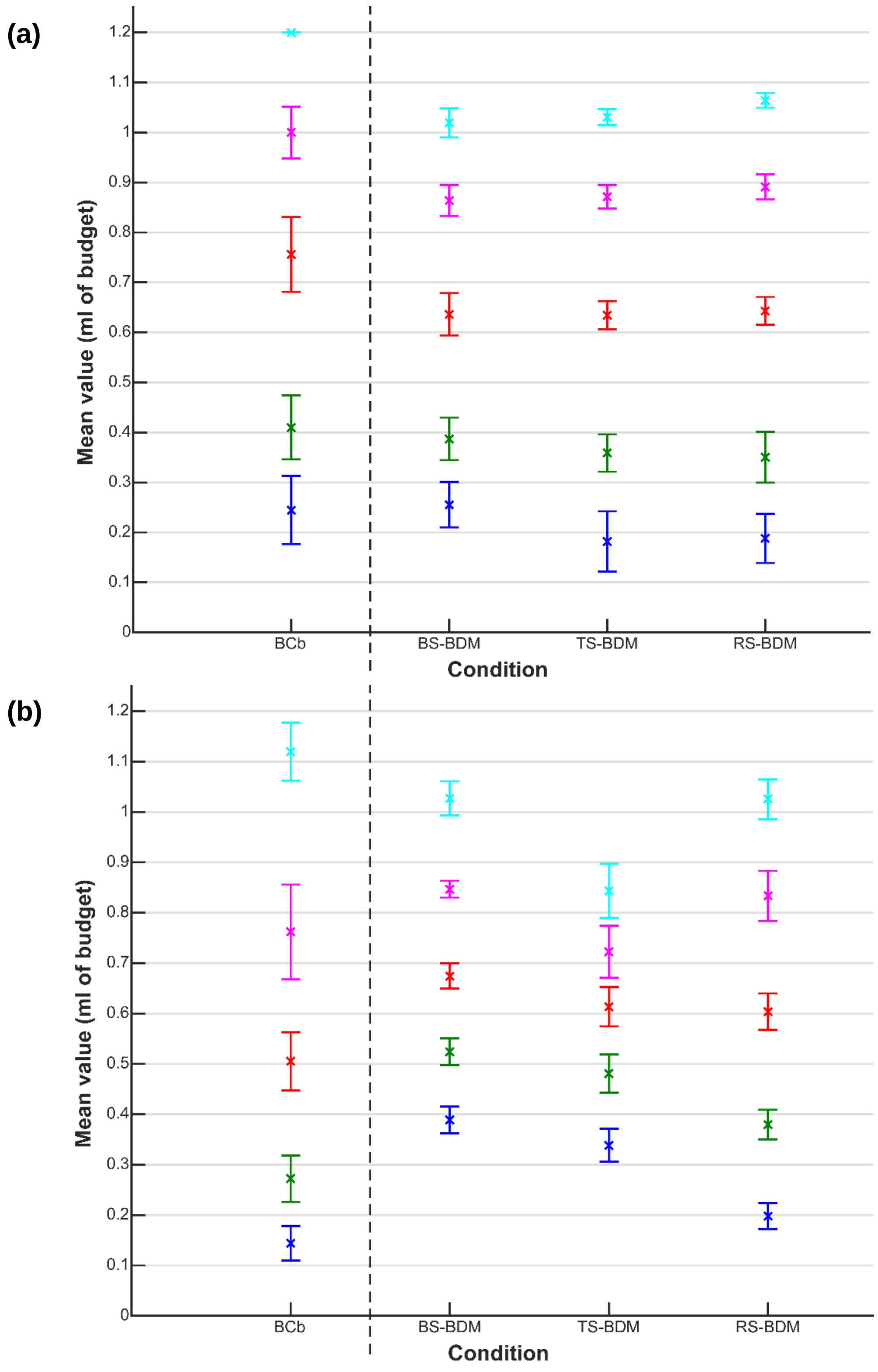


Fig. 3.9a) The values inferred for each reward (0.15ml, blue; 0.30ml, green; 0.45ml, red; 0.60ml, magenta; 0.75ml, cyan) in each of the BDM conditions as well as from the pooled BCb sessions for Ulysses. Error-bars are 95% confidence intervals of the mean. **b)** as in (a) but for Vicer.

Table 3.7) For Ulysses, the p-value is given for the 2-sample t-test between the bids for each reward in each session (S) of the BDM and the values inferred for that reward across all 10 BCb sessions. Sessions for which there is no significant difference ($p > 0.05$) between the two samples for a given reward are shaded in red.

Condition	Reward	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
BS	0.15ml	9.44x10^{-1}	1.97x10^{-1}	1.95x10^{-1}	4.15x10^{-1}	7.40x10^{-2}	1.90x10^{-2}	1.81x10^{-3}	8.05x10^{-1}	1.09x10^{-1}	7.16x10^{-1}
	0.30ml	6.47x10^{-2}	4.34x10^{-2}	2.80x10^{-2}	3.59x10^{-2}	1.64x10^{-1}	2.22x10^{-1}	6.45x10^{-2}	4.62x10^{-1}	8.27x10^{-2}	9.42x10^{-1}
	0.45ml	4.08x10^{-5}	4.04x10^{-2}	2.71x10^{-3}	1.11x10^{-3}	1.22x10^{-2}	6.48x10^{-1}	3.07x10^{-1}	2.87x10^{-4}	1.35x10^{-2}	6.36x10^{-3}
	0.60ml	1.11x10^{-3}	3.61x10^{-5}	3.23x10^{-5}	3.86x10^{-5}	1.70x10^{-4}	2.26x10^{-1}	6.59x10^{-3}	4.05x10^{-3}	1.45x10^{-4}	1.67x10^{-1}
	0.75ml	3.36x10^{-11}	1.68x10^{-18}	2.30x10^{-15}	1.73x10^{-11}	7.25x10^{-18}	1.17x10^{-7}	1.13x10^{-9}	7.22x10^{-14}	6.20x10^{-12}	1.34x10^{-8}
TS	0.15ml	5.13x10^{-1}	5.34x10^{-1}	2.35x10^{-1}	2.93x10^{-1}	1.90x10^{-1}	5.03x10^{-3}	2.90x10^{-4}	3.55x10^{-4}	6.57x10^{-3}	1.20x10^{-3}
	0.30ml	3.24x10^{-1}	1.82x10^{-1}	6.28x10^{-1}	7.22x10^{-1}	7.37x10^{-1}	4.38x10^{-2}	1.67x10^{-1}	2.23x10^{-1}	8.20x10^{-2}	5.40x10^{-3}
	0.45ml	9.45x10^{-3}	1.26x10^{-4}	2.01x10^{-4}	5.27x10^{-3}	6.64x10^{-2}	3.62x10^{-2}	1.07x10^{-1}	1.01x10^{-2}	5.91x10^{-3}	6.47x10^{-3}
	0.60ml	3.07x10^{-3}	2.43x10^{-3}	1.12x10^{-5}	6.33x10^{-3}	5.91x10^{-5}	8.05x10^{-2}	7.86x10^{-4}	3.03x10^{-4}	5.09x10^{-4}	8.70x10^{-3}
	0.75ml	1.60x10^{-14}	6.22x10^{-12}	8.75x10^{-12}	9.82x10^{-13}	1.21x10^{-18}	7.20x10^{-15}	2.60x10^{-16}	4.86x10^{-16}	9.45x10^{-18}	1.31x10^{-17}
RS	0.15ml	8.15x10^{-4}	4.10x10^{-4}	2.33x10^{-2}	4.47x10^{-1}	7.76x10^{-3}	3.49x10^{-1}	8.60x10^{-1}	1.82x10^{-1}	5.04x10^{-1}	5.03x10^{-1}
	0.30ml	1.84x10^{-3}	9.24x10^{-3}	3.73x10^{-2}	8.70x10^{-1}	2.75x10^{-3}	4.94x10^{-2}	5.13x10^{-1}	1.77x10^{-1}	7.46x10^{-1}	8.21x10^{-2}
	0.45ml	2.58x10^{-2}	6.57x10^{-4}	8.42x10^{-3}	5.88x10^{-2}	6.59x10^{-4}	1.02x10^{-2}	4.38x10^{-3}	1.48x10^{-1}	1.02x10^{-1}	6.32x10^{-2}
	0.60ml	8.58x10^{-3}	2.76x10^{-4}	4.08x10^{-2}	4.64x10^{-2}	1.21x10^{-5}	5.02x10^{-3}	2.01x10^{-2}	3.01x10^{-2}	3.82x10^{-3}	4.69x10^{-2}
	0.75ml	5.25x10^{-10}	1.33x10^{-17}	1.84x10^{-10}	5.11x10^{-11}	1.33x10^{-12}	2.06x10^{-8}	2.44x10^{-14}	2.69x10^{-14}	6.01x10^{-15}	2.63x10^{-5}

Table 3.8) For Vicer, the p-value is given for the 2-sample t-test between the bids for each reward in each session (S) of the BDM and the values inferred for that reward across all 10 BCb sessions. Sessions for which there is no significant difference ($p > 0.05$) between the two samples for a given reward are shaded in red.

Condition	Reward	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
BS	0.15ml	1.03×10^{-13}	2.09×10^{-12}	3.48×10^{-14}	3.04×10^{-10}	8.06×10^{-13}	2.35×10^{-12}	7.20×10^{-9}	5.28×10^{-13}	2.84×10^{-10}	8.08×10^{-11}
	0.30ml	3.82×10^{-8}	4.72×10^{-9}	1.74×10^{-10}	1.45×10^{-8}	7.10×10^{-11}	5.54×10^{-8}	7.06×10^{-7}	4.17×10^{-9}	2.95×10^{-6}	2.86×10^{-9}
	0.45ml	3.80×10^{-6}	1.93×10^{-4}	5.80×10^{-5}	7.91×10^{-3}	2.59×10^{-5}	1.20×10^{-2}	1.34×10^{-2}	6.68×10^{-6}	2.38×10^{-3}	2.91×10^{-4}
	0.60ml	3.60×10^{-2}	1.98×10^{-1}	1.59×10^{-1}	4.54×10^{-1}	6.70×10^{-2}	9.78×10^{-2}	1.55×10^{-1}	8.48×10^{-2}	1.63×10^{-1}	7.15×10^{-2}
	0.75ml	1.63×10^{-3}	1.52×10^{-2}	7.06×10^{-4}	9.46×10^{-5}	3.71×10^{-1}	2.49×10^{-1}	1.08×10^{-4}	7.38×10^{-2}	5.92×10^{-2}	1.62×10^{-1}
TS	0.15ml	1.65×10^{-6}	4.74×10^{-6}	7.06×10^{-8}	9.02×10^{-10}	4.49×10^{-14}	6.76×10^{-13}	1.28×10^{-9}	1.19×10^{-9}	6.19×10^{-8}	3.91×10^{-7}
	0.30ml	8.02×10^{-6}	1.36×10^{-3}	5.56×10^{-6}	9.01×10^{-8}	2.87×10^{-10}	2.11×10^{-9}	7.37×10^{-9}	9.95×10^{-8}	4.83×10^{-9}	1.77×10^{-9}
	0.45ml	5.16×10^{-1}	2.10×10^{-1}	2.39×10^{-1}	1.35×10^{-3}	6.03×10^{-5}	4.36×10^{-3}	3.98×10^{-4}	4.33×10^{-5}	5.20×10^{-4}	8.89×10^{-5}
	0.60ml	4.94×10^{-3}	8.60×10^{-2}	9.74×10^{-3}	7.55×10^{-2}	9.81×10^{-1}	9.88×10^{-1}	9.34×10^{-1}	5.89×10^{-1}	5.71×10^{-1}	5.62×10^{-1}
	0.75ml	1.18×10^{-10}	3.70×10^{-10}	2.86×10^{-9}	4.98×10^{-8}	1.43×10^{-6}	1.82×10^{-7}	2.26×10^{-6}	2.14×10^{-6}	6.02×10^{-6}	2.38×10^{-6}
RS	0.15ml	6.14×10^{-2}	7.06×10^{-1}	4.33×10^{-3}	7.05×10^{-4}	2.05×10^{-2}	3.52×10^{-1}	5.02×10^{-1}	1.09×10^{-2}	9.47×10^{-4}	7.07×10^{-3}
	0.30ml	2.32×10^{-3}	4.65×10^{-2}	8.82×10^{-4}	1.39×10^{-3}	1.14×10^{-2}	1.27×10^{-1}	3.14×10^{-1}	1.68×10^{-2}	1.24×10^{-2}	9.54×10^{-3}
	0.45ml	4.78×10^{-3}	1.93×10^{-1}	9.80×10^{-4}	5.38×10^{-2}	2.75×10^{-1}	7.48×10^{-2}	7.85×10^{-1}	1.63×10^{-1}	2.80×10^{-3}	1.73×10^{-2}
	0.60ml	3.32×10^{-1}	7.05×10^{-1}	7.03×10^{-3}	1.81×10^{-3}	7.08×10^{-1}	3.28×10^{-1}	8.13×10^{-1}	2.72×10^{-1}	2.18×10^{-1}	2.19×10^{-1}
	0.75ml	1.41×10^{-3}	4.71×10^{-2}	9.95×10^{-2}	3.11×10^{-1}	6.57×10^{-5}	9.86×10^{-2}	7.52×10^{-4}	2.73×10^{-1}	2.27×10^{-2}	8.17×10^{-2}

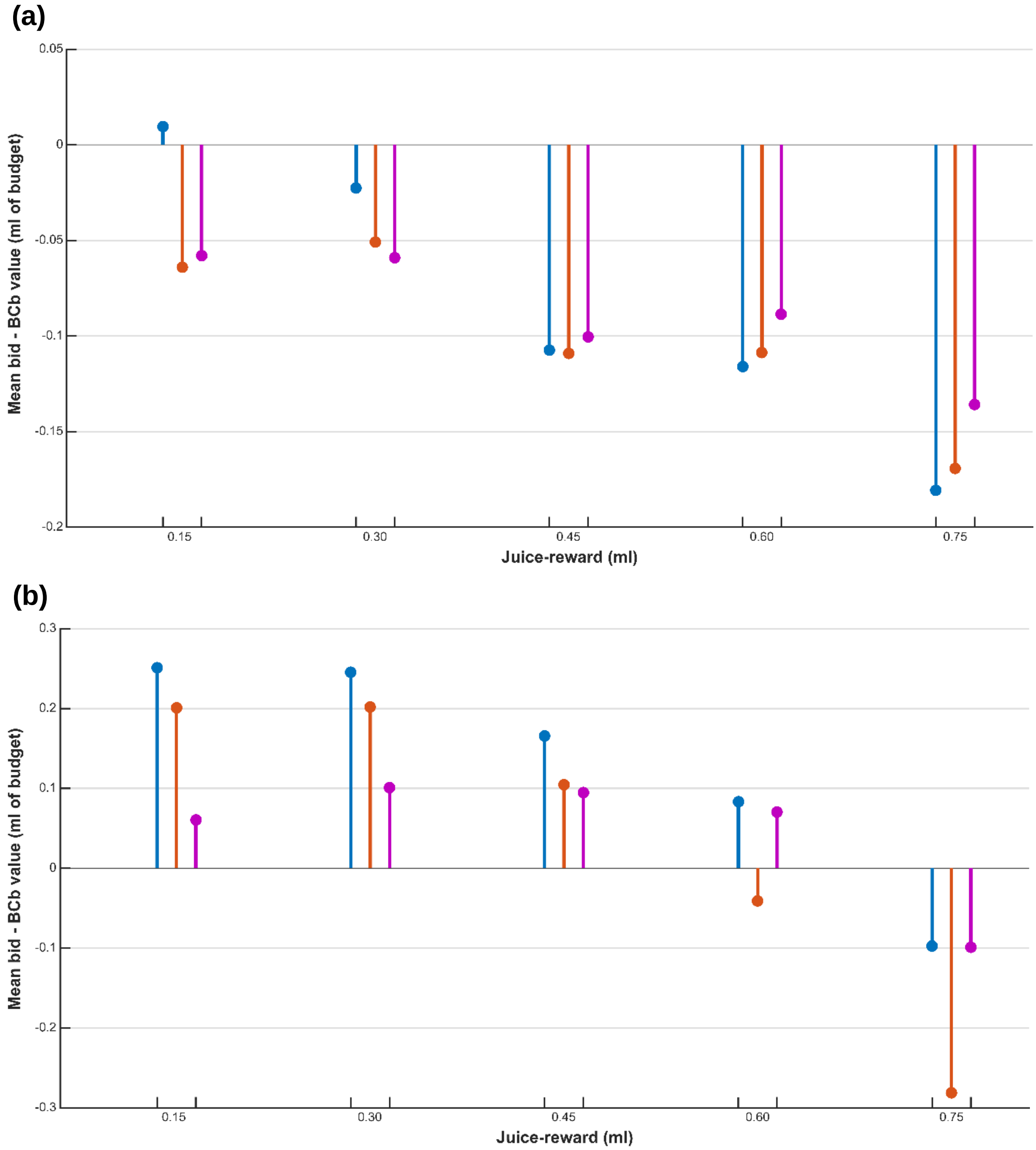
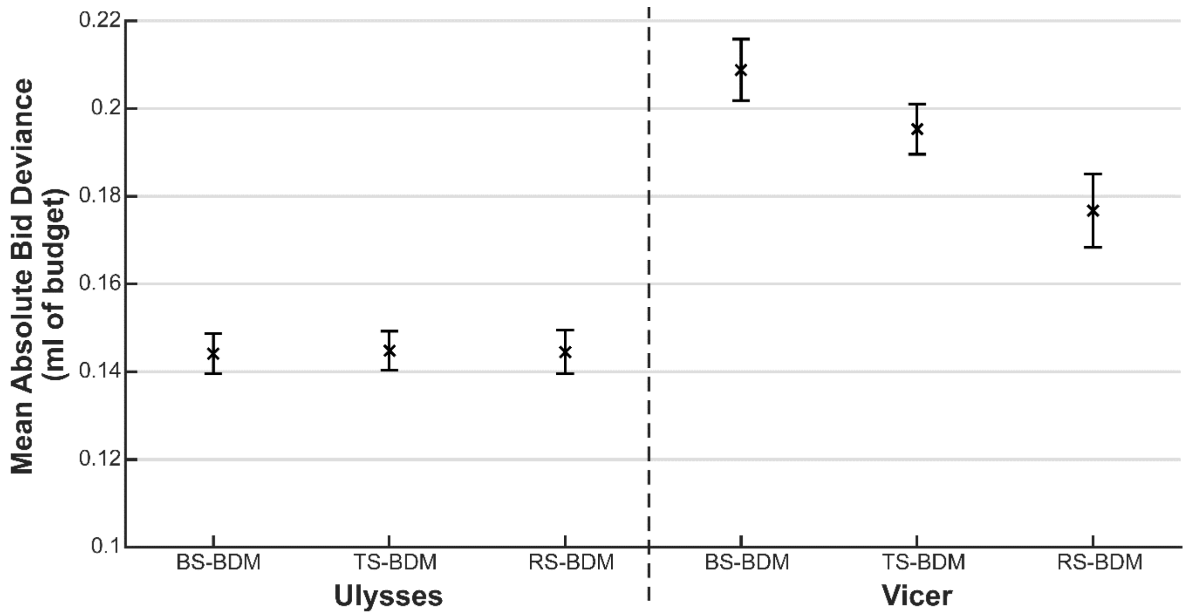


Fig. 3.10a) The difference between Ulysses' mean bid for each juice-reward and the optimal bid, as given by the BCb value, is shown for each of the bid-marker starting position conditions: BS-BDM (blue), TS-BDM (orange), and RS-BDM (purple). Ulysses tended to underbid for all the juice-rewards. **b)** As in (a) but showing a tendency to overbid for most juice-rewards in Vicer. Note that the mean distance of bids from the optimal in the RS-BDM condition was similar for both monkeys.



3.11) The mean absolute bid deviance (ABD) was found for each monkey in each condition by calculating the absolute distance of each bid from the optimal bid implied by their behaviour in the BCb task. Ulysses' mean ABD was the same across all conditions, showing no change over the 30 days of testing. On the other hand, Vicer's mean ABD became significantly small over the course of testing - despite the fact that he had less experience with the random bid-marker start position, which was also more demanding than the BS-BDM and TS-BDM tasks for which the marker's starting position was fixed on each trial. Error bars are 95% confidence intervals of the mean.

The lack of any difference in the magnitude of Ulysses' mean ABDs across the different conditions makes it tempting to conclude that there was no improvement in his bidding over this testing period. However, while the mean ABDs give an impression of the precision of bidding behaviour in terms of the magnitude of the bid, the payoffs that the monkeys experienced on each trial were more likely to be driving their behaviour. The monkeys had no reference frame within which 'optimality' was meaningful, apart from their history of payoffs for each bid - they had to learn through trial and error.

Therefore, the optimality of the monkeys' performance may be better evaluated in terms of their payoffs. As such, the payoffs of an optimal bid facing the same computer bids as the monkeys were calculated for every trial. For each trial the costs incurred were found by finding the difference between the payoff of an optimal bid and the monkeys' actual payoff. The mean cost per trial was then calculated for each BDM condition (fig. 3.12).

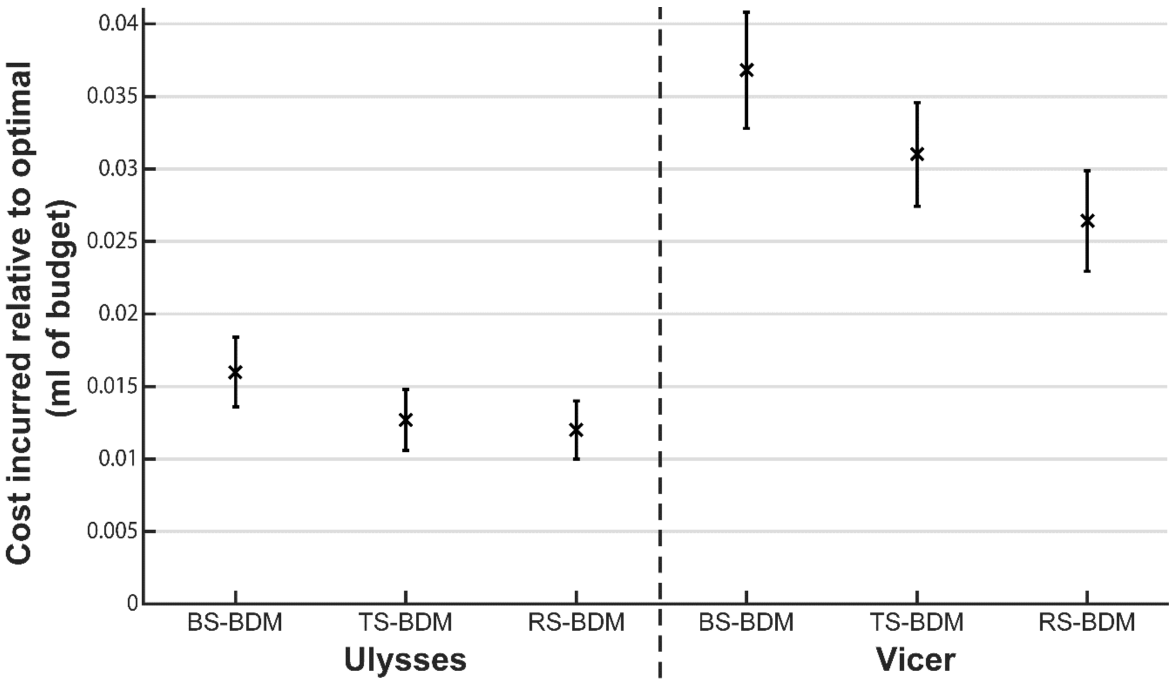
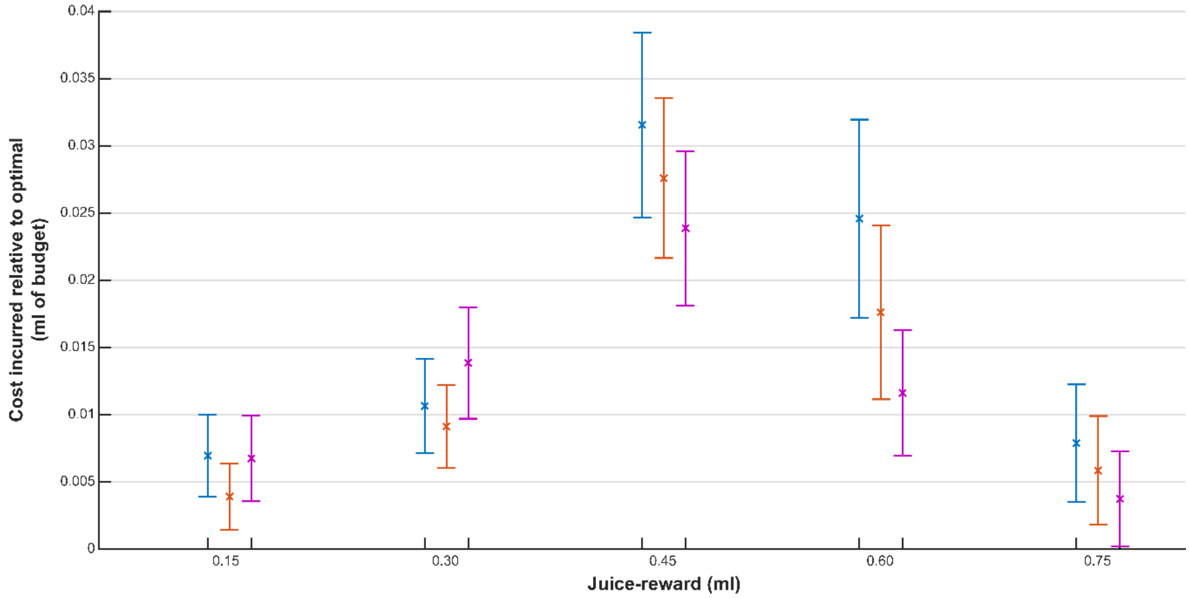


Fig. 3.12) Mean per trial costs incurred were found for each monkey in each condition by calculating the difference between the monkey's actual payoff on each trial and what their payoff would have been had they placed an optimal bid. Both monkeys showed an improvement in terms of their per-trial payoffs over the course of BDM testing, incurring significantly lower per-trial costs in the RS-BDM than in the BS-BDM. The theoretical expected cost of misbehaviour (ECM) and the actual incurred costs relative to optimal bidding were not found to be significantly different for any of the juice-rewards, indicating that the intended and realised payoff schedules were similar. Error-bars are 95% confidence intervals of the mean.

While it is possible that Ulysses was approaching an asymptotic level of performance by this stage of his training in the BDM task, he nevertheless did continue to improve his bidding in terms of the costs incurred on a trial-by-trial basis: Bonferroni-corrected multiple-comparison t-tests found a significantly lower per-trial cost in the RS-BDM than in the BS-BDM for both monkeys ($p < 0.05$), though differences in incurred costs between the TS-BDM and the other two conditions were not statistically significant. Notably, the mean per trial costs incurred by Ulysses were significantly lower than those in a preceding version of the task that was used to train the monkeys (Ch. 5.6, fig. 5.30b), suggesting continued improvement in his performance.

The costs incurred on every trial for individual juice-rewards (fig. 3.13) reflected the payoff schedule that was imposed by using a Beta (4,4) distribution for the random generation of computer bids (fig. 3.14), with greater costs for the same ABDs for juice-rewards whose true values were closest to the centre of the budget-range (fig. 3.15).

(a)



(b)

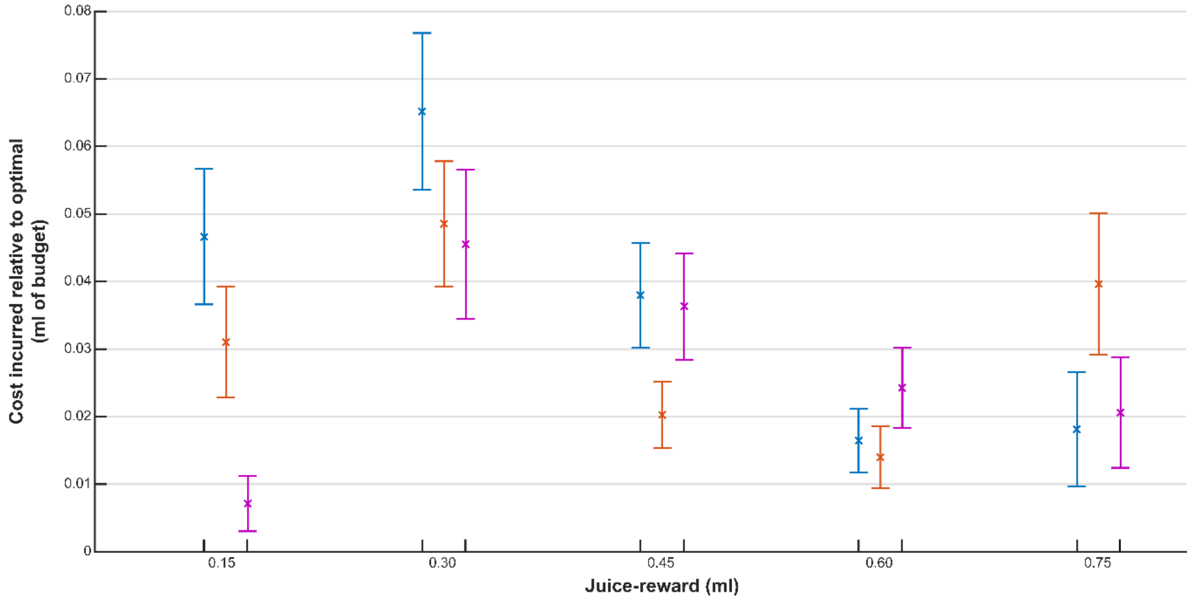


Fig. 3.13a) Ulysses' mean cost incurred per trial when bidding for each juice-reward in the BS-BDM (blue), TS-BDM (orange) and RS-BDM (purple). Error-bars are 95% confidence intervals of the mean. **b)** as for (a) but for Vicer.

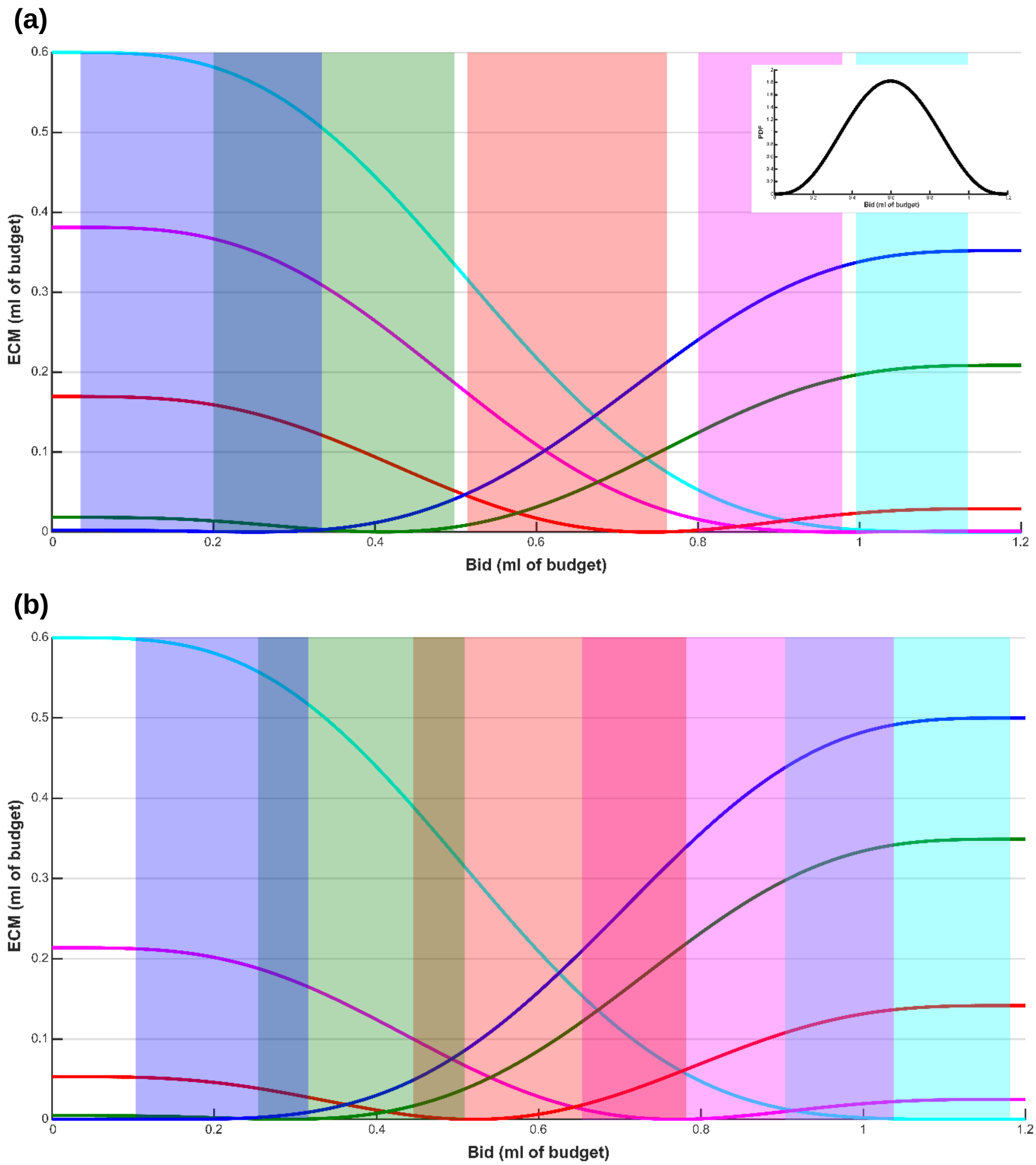


Fig. 3.14a) The expected cost of misbehaviour (ECM) at different bids for each of the 5 juice-rewards (0.15ml, blue; 0.30ml, green; 0.45ml, red; 0.60ml, magenta; 0.75ml, cyan) given Ulysses' values for those rewards. The inter-quartile range of bids for each reward in the RS-BDM is shown overlain as a transparent rectangle in the same colour as the relevant ECM line. Bids were clustered in regions of low expected cost. The Beta (4,4) distribution is inset for reference.
b) As for (a) but for the second monkey, Vicer.

(a)

(b)

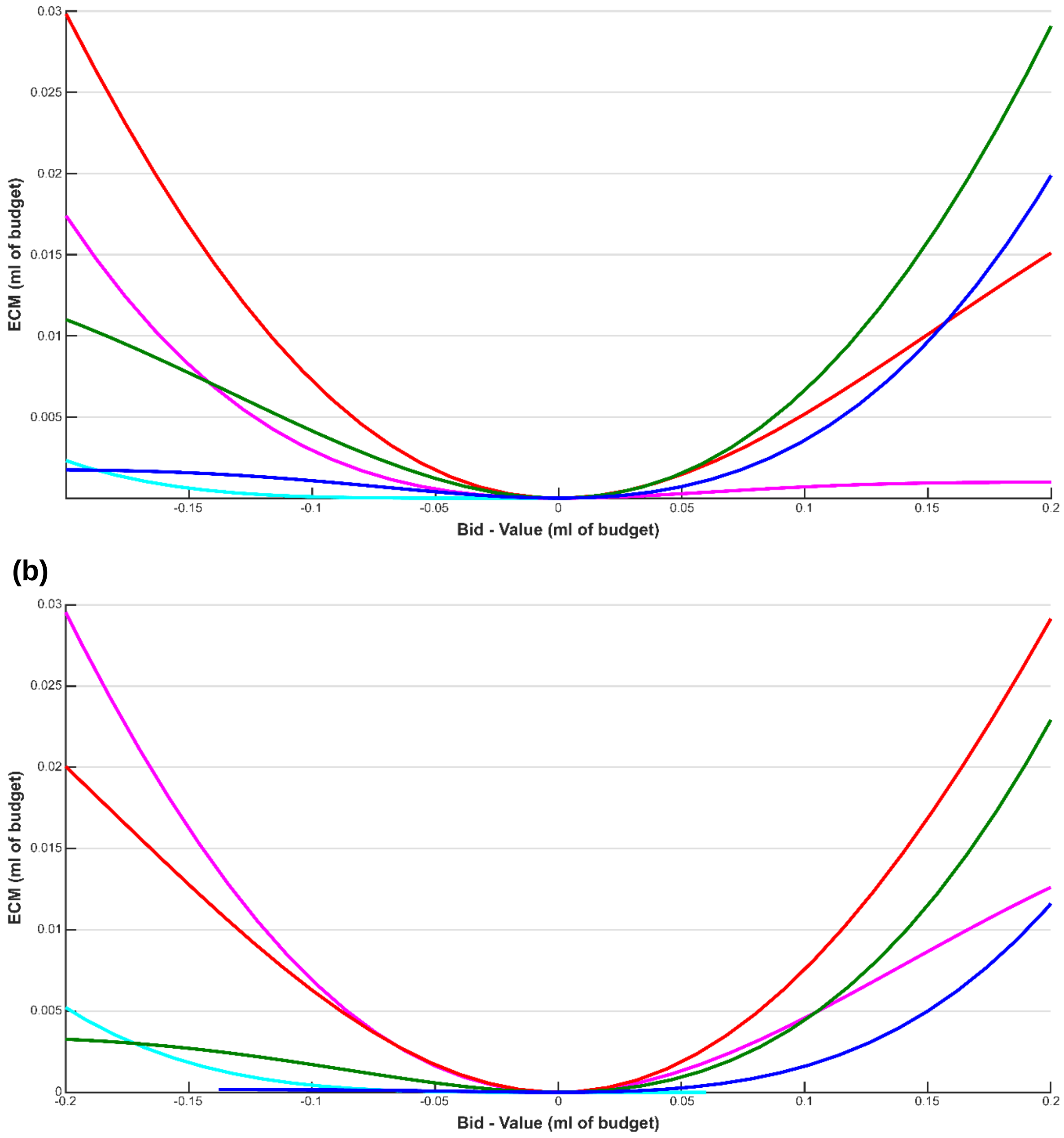


Fig 3.15a) For Ulysses, the ECM is expressed for each juice-reward (0.15ml, blue; 0.30ml, green; 0.45ml, red; 0.60ml, magenta; 0.75ml, cyan) in terms of the difference between each bid and the monkey's value for that reward. The ECM for the same absolute deviations in bidding was highest for those rewards whose values were closest to the centre of the bidding range, where the peak of the Beta (4,4) distribution is located. **b)** Showing the ECM as in (a) but for Vicer.

Note that the costs shown for the 0.75ml reward for Ulysses were an underestimate of the true costs that he experienced, as they made use of the optimal bid value of 1.2ml, whereas his true subjective value for the reward was some unknown amount greater than the water-budget maximum of 1.2ml (fig. 3.8a). Nevertheless, estimating the costs by using some implausibly high value for the 0.75ml juice-reward (e.g. 2ml of water-budget) does not change the trend of improvement seen between the BS-BDM and the RS-BDM: costs incurred are still significantly lower in the RS-BDM in that case ($p = 0.007$). Moreover, the pattern of relative differences in costs incurred for the different juice-rewards in the RS-BDM also don't change dramatically: costs incurred for the 0.75ml juice-reward remain lower than those for the 0.45ml juice-reward and become comparable to the costs incurred during bidding for the 0.3ml reward (~0.014ml of water-budget per trial).

The costs incurred per trial for individual rewards might help to explain the differences in the distance from the optimal bid between the high and low valued juice-rewards (fig. 3.10). For example, for Ulysses in the RS-BDM, the mean bid for the 0.6ml juice-reward was 0.09ml less than the optimal bid, whereas the mean bid for the 0.3ml reward was only 0.06ml less than the optimal. However, the mean per-trial costs of these bids were equivalent, despite their different distances from the optimal bid - this difference in costs for the same ABDs in different juice-rewards can be seen more clearly in figure 3.15.

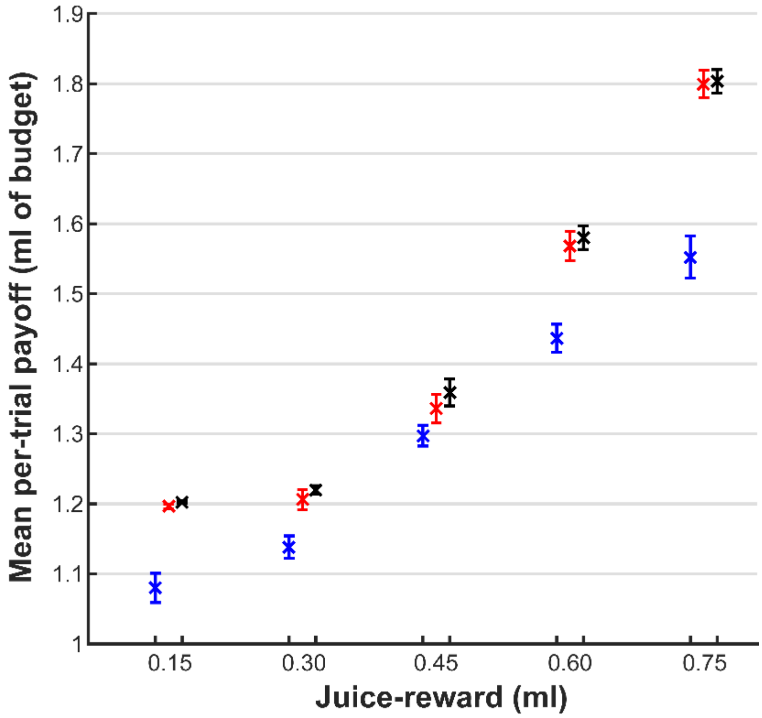
These results are consistent with the idea that the differences in payoffs - and therefore the different costs - drive learning of the task, as the different distances from the optimal bid may reflect a degree of cost that the monkey either cannot detect or, simply tolerates - this possibility is discussed further in the next section (Ch. 3.3).

While it is clear that the monkeys' mean bids are not equal to their values as inferred from the BCb task, they nevertheless seem to reflect their order of preference for the different juice-rewards. Moreover, their bids continue to approach optimality as they acquire more experience in the BDM. Given this, and the failure of human subjects to bid in a precisely optimal way, it is perhaps more useful to consider the monkeys' behaviour in terms of a continuum between totally random behaviour and optimal behaviour. To do this, we compared the mean per-trial payoff of an optimal bidder, the actual payoffs of the monkeys, and, a simulated ‘random’ bidder who randomly selected a bid on each trial from a uniform distribution with support [0, 1.2] (fig. 3.16).

This comparison shows the extent to which the monkeys’ behaviour had approached optimality in terms of their payoffs, and supports the suggestion that the shallow payoff schedule of the BDM limits the incentives to bid optimally, as the monkeys’ average per-trial payoffs were close to optimal for most of the juice-rewards despite an absolute deviation from the optimal bid of ~0.14ml and ~0.18ml in Ulysses and Vicer respectively.

Considered in these terms, both monkeys performed much more similarly to an optimal bidder than a ‘random’ one, and their deviations from the optimal bid may have reflected a limit with regards to the ability of the BDMs payoff schedule to motivate more precise bidding.

(a)



(b)

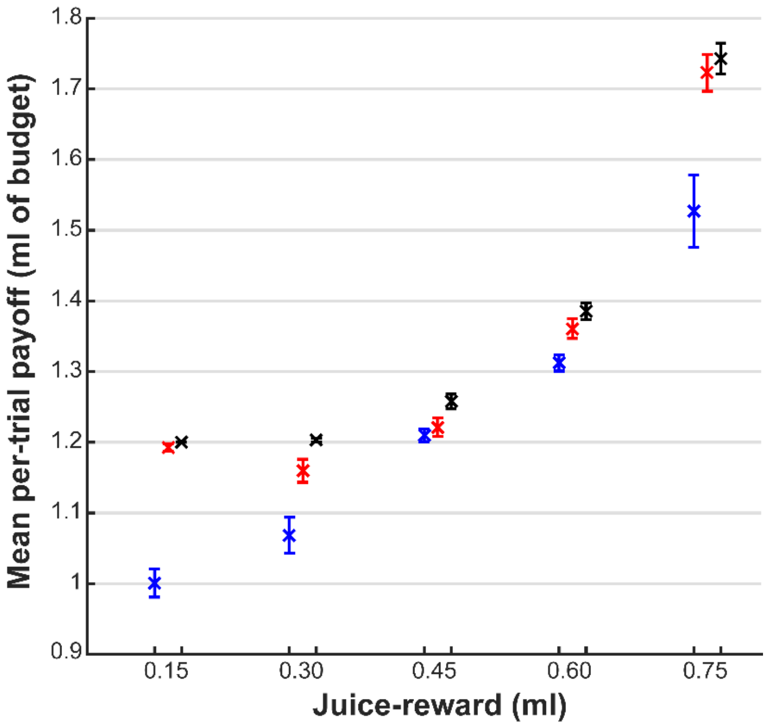


Fig. 3.16a) Mean per trial payoffs over 10 sessions of 200 trials each for an optimal bidder (black) and a random bidder bidding according to a uniform distribution (blue), as well as the monkey's actual per-trial payoff from the 10 RS-BDM sessions (red). The optimal and random bidders both faced the same computer bids as the monkey did. Error-bars are 95% confidence intervals of the mean. **b)** The same simulation as in (a) but using data from Vicer.

3.3 - Discussion of BDM results

The results of the BDM in monkeys show that they can learn to place bids that reflect their preferences, and can differentiate 5 different reward objects in a single session of testing. Their ability to do this regardless of the motor contingencies of bid placement suggests that their learning of the task does not depend on the reproduction of fixed stimulus-response associations, but rather implies learning of the significance of different bids in terms of the expected payoffs associated with them.

Ulysses had approximately one more year of BDM training than Vicer and this was reflected in his superior task-performance. For Ulysses, the bids for all 5 juice-rewards could be distinguished by statistical methods in each of the 30 BDM testing sessions and on average this could be achieved within ~113 trials. On the other hand, Vicer’s bids could always be used to differentiate at least 4 of the 5 rewards, but only successfully differentiated all juice-rewards in 21 of the 30 sessions. Furthermore, even in those sessions in which his bids did reliably distinguish all 5 rewards, he took significantly more trials than Ulysses to do so, requiring an average of ~180 trials - though both monkeys always completed the same number of correct trials (200) per session.

Thus, at Ulysses’ level of performance, significantly fewer BDM trials could have been used to acquire estimates of his values and reliably rank 5 rewards - which would represent and advantage of the binary choice method for the purposes of neurophysiological recording. On the other hand, Vicer required almost all of the 200 BDM trials in a given session in to achieve this, making the task comparable to the BCb task in terms of the speed with which rewards are reliably differentiated. However, other performance indicators, such as mean per-trial costs incurred relative to optimal (fig. 3.12) suggest that Vicer was still learning and had not yet achieved a stable level of performance.

Taken together, this suggests that Vicer’s bidding was less consistent than Ulysses overall. This difference in the monkeys’ performances was reflected both in the proportion of their bid variance explained by the identity of the juice-rewards, and the magnitude of Spearman’s Rho for the correlation between their bids and the ordinal rankings of the juice-rewards (Ulysses: Mean = 0.91, SD = 0.017; Vicer: Mean = 0.81, SD = 0.053). While Vicer's bidding continued to improve over the course of these sessions, this yielded another problem insofar as it made it difficult to distinguish the effects of the various conditions that were tested from the effects of learning over time.

Ideally, both monkeys would have been tested in this task at a later stage, when they had reached a more stable level of performance, but time limitations prevented this. To a large extent Ulysses had reached a stable level of performance, with only minor improvements in the optimality of his bids being observed over the course of the 30 sessions presented here. However, Vicer's performance was notably worse, with considerably more error trials, though this was mostly due to many consecutive trials that were completely ignored (usually towards the end of a testing session), rather than an inability to perform the task. Vicer was a younger monkey and was far more distractible in general, but his far higher error rate (~31% of all trials compared to ~7% for Ulysses) may have been indicative of a general lack of motivation, and therefore inattentiveness, which may have impacted upon the precision of his bids.

Nevertheless, the bidding of both monkeys was consistent enough to conclude that the BDM could be used to rank rewards in terms of their order of preference, and as Vicer could do this reliably in every session of a previous version of the BDM task which made use of only 3 juice-rewards (Ch. 5.6), it is likely that the difficulties that he encountered in consistently differentiating 5 juice-rewards could be counteracted with more extensive training.

More difficult to assess was the optimality of the monkeys' bids. As the optimal bid is equal to a monkey's value for the reward, the optimality of the monkeys' bids acts as an indication of how reliable their mean bids were as an estimate of their reward values. This was measured by comparing the monkeys' mean bids with their values for the juice-rewards as inferred from a supplementary binary choice (BCb) task, which acted as a benchmark, as the only incentive compatible method of subjective value elicitation in animal subjects.

The mean bids of both monkeys were significantly different to their BCb values for most of the juice-rewards that were tested. Both monkeys therefore failed to reproduce truthful reports of their subjective values by bidding in the BDM - only their correlated underlying preferences were reliably revealed.

It is unclear whether the BDM failed to elicit the monkeys' true subjective values due to a lack of understanding of the task, non-incentive compatibility[54], or their cognitive limitations. A definitive answer to this question requires further experimentation, however, in our opinion the results presented here are consistent with the latter possibility and suggest that it is unlikely that the monkeys simply did not understand the task contingencies.

First, the monkeys' bids were in accordance with their preferences and did bear some relation to their subjective values as they clustered in regions of the bidding-range in which the costs of sub-optimal bidding were low. Second, the monkeys' bids approached their optimal value as they acquired greater experience in the BDM. This was reflected in the absolute distances between their bids and the optimal bids for the different juice-rewards, as well as in the progressively lower costs that they incurred in the conditions that were tested later. Third, the mean costs incurred on a given trial were both small in absolute terms and relative to the payoffs of an optimal bid. In fact, it was not certain that the monkeys could even discriminate costs of such magnitude (and specific tests would be required to confirm that they can). Indeed, the costs that the monkeys incurred on a per-trial basis were in the range of variability that is observed in the solenoid juice-delivery systems (Appendix 2).

Moreover, even if the monkeys could discriminate such small differences in payoffs, it is still possible that the motor and cognitive effort costs associated with precisely determining and placing an optimal bid outweigh the benefits[50] - in that case, following a noisier strategy may even have a greater net utility, especially as we ensured that the monkeys were always sated on liquid by the end of each day.

Therefore, in the strictest sense the monkeys did not place bids that were equal to their values for the different juice-rewards, but nor did they behave as if they did not understand the task contingencies. The third possibility, that the BDM is simply not incentive compatible, cannot be ruled out but requires extensive testing that is beyond the scope of these experiments. In principle, if it could be shown that the monkeys' behaviour is better captured by a utility model for which the BDM is not incentive compatible, then that could explain the disparities between their BCb and

BDM values. But few experiments in monkeys have gone beyond a characterisation of their behaviour in terms of a Von-Neumann-Morgenstern Expected Utility model.

With regards to concerns about the underlying assumptions that are made of the monkeys' preferences, and more generally when considering the BDM as a method of value elicitation, the words of George Box seem relevant: "Essentially, all models are wrong, but some are useful"[61].

Taking a more practical approach to these results, it is apparent that the monkeys can use the BDM to produce bids that reflect their preferences on an analogue scale. Importantly, they were capable of distinguishing 5 differently valued rewards and therefore 5 distinct levels of utility, which is a higher resolution than is achieved in some studies that use the BDM with human subjects. For example, Plassman et al. (2007) made use of only 4 possible discrete bids[14] and could therefore only correlate neuronal activity with 4 different levels of their subjects' utilities - note, they did not measure utility directly, but distinct subjective values will necessarily be at a different utility levels.

Furthermore, none of the neuroimaging studies that have used the BDM in human subjects have verified the incentive compatibility of the method, or the optimality of subjects' bids. Rather, the BDM has simply been used as a fast and easy way of acquiring trial-by-trial responses that should correlate with the subjects' values and utilities. This is adequate for many applications, and can be used to identify brain regions involved in decision making and the representation of value, despite any noise in the subjects' responses, or, the potential and uninterrogated deviations from their true values.

Human subjects also do not bid perfectly in the BDM, and their deviations from optimality are especially pronounced when they are not informed of the optimal strategy[37]. Subjects in the minimal information condition of Irwin et al.'s (1998) second induced value experiment (Ch. 2.1) submitted bids that were between ~$1 and ~$1.25 away from the optimal bid - a deviation of between ~16.7% and ~20.8% of the total bidding range of $6. Even subjects in the full information condition made bids that were ~$0.25 and ~$0.5 away from the optimal bid - a deviation of between ~4.2% and ~8.3% of the total bidding range. In comparison, Ulysses' bids in the RS-BDM task deviated by ~0.14ml, or ~11.7% of the total bidding range, while Vicer's bids in that condition deviated by ~0.18ml, or ~15% of the total bidding range (fig. 3.11).

Thus, the optimality of the monkeys' bids in the BDM task developed here is similar to that of human subjects, and neither species has been found to consistently place optimal bids when tested with this method. It is perhaps more practical to consider optimality as a continuum between completely random behaviour and optimal behaviour (fig. 3.16) in the context of an individual subject's decision making - wherein optimality is a matter of degree.

Unless we are to expect more of monkeys than of human subjects (and perhaps such a case can be made), then by the standards applied to human studies of decision making the BDM method can be used to elicit a monkey's subjective values.

4

Training a monkey in the BDM

The BDM is a more complicated task than the binary-choice (BC) task, which is the current standard for truthful value elicitation in animal experiments. The various elements and contingencies that must be integrated to identify an optimal decision are far more numerous and complex than those required for correct performance in a binary-choice task. And yet, the principles of reinforcement learning should deter the sceptic from assuming that monkeys cannot learn the BDM.

If there exists a perceptible gradient in net payoffs, then action weights should be adjusted by reinforcement (in a manner proportional to the magnitude of that gradient) such that the payoff-maximising action becomes preferred. To restate this, when the subject has enough time and experience with the payoffs and contingencies of the BDM, they need not understand the reasoning or mathematics that defines the optimal bid as their best choice, rather, they can search for this optimum by reinforcing actions in a manner that is proportional to their expected payoffs.

There is much evidence to suggest that macaques and other animals do learn to maximise rewards as described by theories of reinforcement learning[34]. Considering this, the BDM task may come to be viewed as a continuum armed bandit-task: a generalisation of the more commonly tested *n*-armed bandit tasks that have been extensively used with monkey subjects in the field of decision neuroscience[36]. To conceptualise this, let us consider a 'discrete' BDM task in which only ***n*** possible bids can be made. In such a situation, the animal will come to have an expectation of the value of the probabilistic outcomes associated with each of the ***n*** bids – much as each arm of the *n*-armed bandit will come to have an associated expected value which determines the rate at which it is selected.

In the *n*-armed bandit task, the animal doesn't need to understand the contingencies that lead to the probabilistic reward, only the overall payoffs of each action. Similarly, in the BDM, the monkey doesn't need to know why a given bid is best, only that it has the highest expected payoff. Thus, there is no inherent feature of the BDM that should prevent the monkey from finding the best bid. Rather, the key obstacle to learning is the large number of trials that are required to form a reliable expectation of the payoffs of each action/bid.

Apart from that, the BDM is learnt like any other task: first, the animal is taught what each stimulus signifies in terms of reward type, magnitude and timing; second, the animal is taught how to provide a behavioural response in the task, and how the outcomes are contingent on their response – simply through exposure; third, the subject participates in the task itself, learning through experience. The first two of these stages of learning are described in the first of the two sub-sections that complete this chapter, whilst the second section addresses the results of the monkey's initial training in a preliminary version of the BDM task.

All the training experiments and preliminary tests presented in this chapter made use of the same experimental set-up as was used in the final version of the BDM – see Appendix 2 for a description of these experimental methods.

All the data presented in this chapter are from the first monkey, Ulysses, and are used to outline the rationale underlying the design of the training tasks and to illustrate their effectiveness in allowing the monkey to learn about the stimuli used in the final BDM task as well as how to use a joystick to place bids. The second monkey, Vicer, was trained using the same tasks once these had been established in Ulysses.

1 – Teaching the components of the BDM

In this section, we present the training paradigms used in the first monkey to achieve meaningful performance in the earliest ‘preliminary’ version of the BDM task. Broadly, there were two types of training we conducted before the preliminary BDM: Pavlovian Stimulus Learning (PSL) of the stimuli that would be used to signify the reward and budget in the BDM, and, operant conditioning of joystick control, providing acquisition of a controlled movement with which the monkey could make their bids. All the training tasks described in this section make use of the same experimental set-up, including monitor position, joystick, juice-delivery system etc. as described in the methods section of the final BDM task (Ch. 2.2).

The training stages for the BDM task progressed with a primary focus on finding out whether the method could work at all. In other words, we first wanted to establish whether a monkey could make meaningful bids at all in a basic functional version of the task before we invested time in secondary tasks that would be required for a more detailed analysis. As such, our goal was to quickly achieve a level of performance in the task that could be subjected to meaningful analysis, with the intention of returning to previous stages of training and refining the task if we had sufficient evidence to deem this a worthy use of time, resources, and an animal subject.

If we felt confident that a certain stimulus or contingency had become well learnt then we did not delay the process of moving on to the next stage of training. But this very time-sensitive and coarse strategy had its fair share of problems. At times, it has meant that the training data are incomplete, or that certain aspects of training were rushed, or, that several parameters were changed at once; meaning that later comparison between sets of training tasks was often made more difficult.

Perhaps the most problematic aspect of this approach, however, was the fact that the training of the first monkey in a binary choice task was significantly delayed. This made it difficult to assess early bidding behaviour beyond a simple ordering by preference, and complicated the process of selecting rewards that made good use of the monkey’s bidding range.

Nevertheless, it also had its benefits: early in the monkey's training we were able to establish that the BDM task was not only feasible in principle, but that monkey's bids could be used to differentiate between reward objects. It also meant that we had enough time to return to only the most relevant stages of training and improve the most pressing elements of the monkey's performance in the task (notably joystick control), as well as refine the BDM task itself through changes to its structure and the presentation of stimuli. And once these aspects of the BDM and the training leading to it had been worked out in the first monkey, the training of the second monkey proceeded in a far more systematic manner – with more time spent on ensuring good performance in the key training tasks and complementary tests, such as the binary choice task, before progressing onto a refined version of the BDM.

Pavlovian Stimulus Learning of the juice-rewards:

After animals were habituated to the experimental chamber they immediately took part in daily experimental sessions designed to teach them the association between different volumes of juice reward and unique fractal images. Fractals were used to signify rewards as their lack of occurrence in the natural world means that they are completely novel stimuli for the monkeys, with presumably no previous associations or learnt values from the animal's experience. Moreover, fractal stimuli are well suited for neuronal recordings - their inconsistent visual features are less likely to systematically excite neurons in recording sites of interest, e.g. Orbitofrontal cortex or midbrain DA neurons (see Ch. 6.1 for recording sites of interest).

PSL for the reward to be bid for proceeded using a simple task design (fig. 4.1). Following a variable inter-trial delay, a cross was presented to orient the subject and capture their attention. The reward-predicting fractal was then presented, and after a 1.5s delay the associated volume of juice reward was delivered to the monkey. Initial training sessions used only 2 or 3 different fractals/reward volumes, with more stimuli/rewards being added in later sessions. Stimulus/reward presentation was randomised to maximise learning rate by making use of the attention enhancing effects of prediction errors[62,63,64].

Fig. 4.1) Trials start with a 0.5s preparation period to orient the subject to the screen and the upcoming reward-predicting stimulus. A fractal image is then shown, indicating the delivery of a specific volume of juice reward. The appearance of a red border signals the delivery of reward in 0.5s. Reward delivery was followed by a variable inter-trial interval (ITI) of 2±1s, with random durations distributed by a truncated exponential function.

PSL trials were designed to be as consistent as possible with the BDM task itself. Hence, both the size of the fractals and their location on-screen were the same in PSL trials as would later be used in the BDM, with the fractal image being slightly lateralised in both cases.

However, in the BDM task, the timing between the first presentation of the reward stimulus and its receipt by the animal is necessarily delayed by the process of bidding, as well as by other epochs relating to stimulus changes that are not present during PSL sessions (BDM task shown in fig. 4.10), and using such long delays between stimulus presentation and reward receipt during PSL would slow the acquisition of the CS-US association, either through a direct effect of the CS-US interval according to trial-based learning models[65], or, through a reduction of the cumulative reinforcement rate in rate-based models[66]. To further complicate matters, the subjective value of rewards is influenced by temporal discounting[5], with delays leading to reductions in the subject's value for a reward[67,68].

Therefore, a second stimulus was used in both BDM and PSL trials, with the appearance of a red border around the fractal image indicating that the associated reward would be delivered in 0.5s. It is this delay that was made to be consistent between the two conditions; allowing us to use shorter intervals between the first presentation of the fractal image and reward receipt in PSL sessions than in the BDM.

We wanted to determine whether the subject had learnt the different values indicated by the fractals before starting the BDM – to know if we could reasonably expect the subject to differentiate between rewards with their bids, while also wanting to start BDM training as soon as possible to see whether it was at all feasible[*].

In the absence of a choice-task it is difficult to ascertain whether learning of the stimulus value has taken place. However, other proxy measures for reward-value can be recorded, such as the licking rate prior to reward receipt: It has been previously demonstrated that monkeys will lick a juice spout with greater frequency as reward expectation increases[69,70].

Using an infrared sensor positioned 0.2cm below the juice-spout, we inferred the licking rate from the number of times that the infrared beam was interrupted by the monkey's tongue in the period preceding reward receipt. As our subject was not head-restrained this measure was not highly reliable; for example, the monkey could move their entire mouth over the spout if they wanted to, and interrupt the beam for the duration of the trial, or, conversely, position themselves such that they could lick the spout without interrupting the beam – this is particularly problematic for comparison between sessions, where the animal may assume different positions on different days.

Nevertheless, we were able to record licking in the first 4 of the 12 fractal PSL sessions and used this as an approximate indicator of whether or not learning had taken place. The average normalised number of licks at a given time point for a given reward-predicting stimulus was calculated by summing all recorded beam-interruptions from all trials in each 0.02s time-bin (corresponding to the 50Hz frequency of data collection from the infrared sensor), and then dividing this total number of licks by the total number of trials on which that reward predicting stimulus was presented. The frequency of licking at a given time point is then calculated by

[*] For the second monkey, Vicer, we did not have to do this as we conducted binary choice testing to infer the values of stimuli before beginning the BDM; we could afford to invest time in binary choice testing first having already established the feasibility of the BDM task.

dividing this average number of licks per trial per bin by the length of time covered by each bin.

Thus, for each reward, where **nT** is the total number of trials for that reward, the average number of licks per trial for a given bin are given by:

$$\boldsymbol{Licks\ per\ Trial} = \frac{\sum_{T=1}^{nT} \boldsymbol{Beam\ interruptions}}{\boldsymbol{nT}}$$

The lick frequency at a given time point or bin is then given by:

$$\boldsymbol{Lick\ frequency} = \frac{\boldsymbol{Licks\ per\ trial}}{\mathbf{0.02}}$$

The data were consistent with the animal having learnt both the volume and timing (fig. 4.2-3) of reward predicted by the fractal stimuli and the appearance of the red border at the start of the ‘Pre-delivery’ epoch.

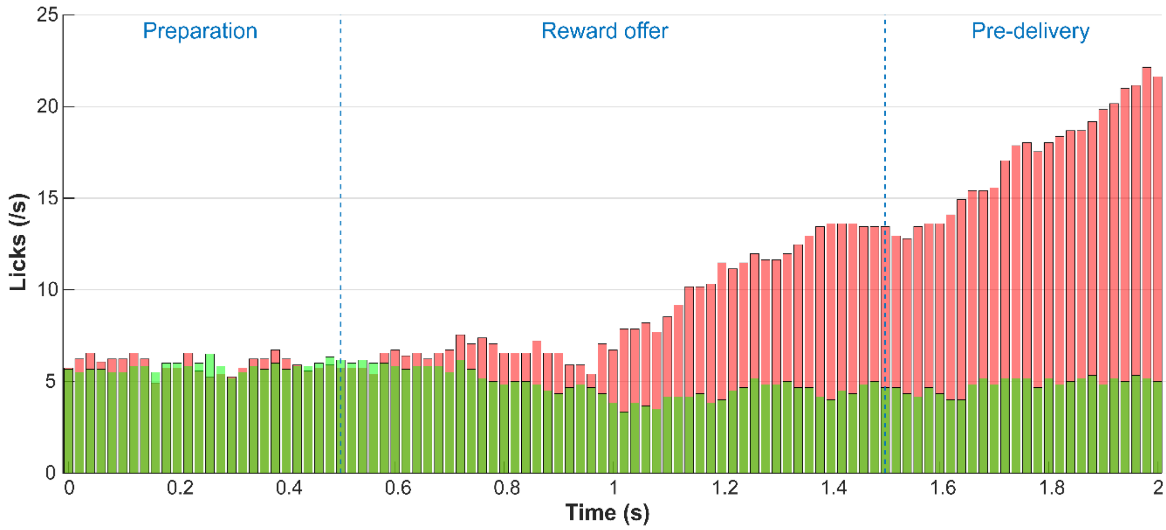


Fig. 4.2) The average normalised frequency of licking sampled in an example PSL session in which 0.8ml (red) and 0ml (green) juice-reward predicting fractals were used. Epoch boundaries are shown by the dashed blue lines.

Inspection of the lick frequency in figure 4.2 reveals that before stimulus presentation there is no difference in the frequency of licking between 0.8ml and 0ml trials. An increase in lick frequency can be observed after presentation of the 0.8ml predicting fractal, whilst a decrease in lick frequency is seen following presentation of the 0ml predicting fractal. Note also the increased rate of licking in anticipation of imminent juice delivery following presentation of the red-border in the 'Pre-delivery' epoch for the 0.8ml predicting fractal. These observations were confirmed by a 2-way ANOVA with factors of trial 'Epoch' (3 levels) and reward 'Volume' (2 levels), finding significant main effects for both trial epoch, [$F(2,196) = 125.75$, $p = 7.32 \times 10^{-36}$], and reward volume, [$F(1,196) = 401.16$, $p = 2.66 \times 10^{-49}$], as well as a significant interaction between these two factors, [$F(2,196) = 155.12$, $p = 4.11 \times 10^{-41}$].

This interaction between on-screen stimulus changes, captured by the 'Epoch' factor, and the anticipated reward volume can be seen more clearly in figure 4.3, where the mean number of licks for each epoch in this example session are presented.

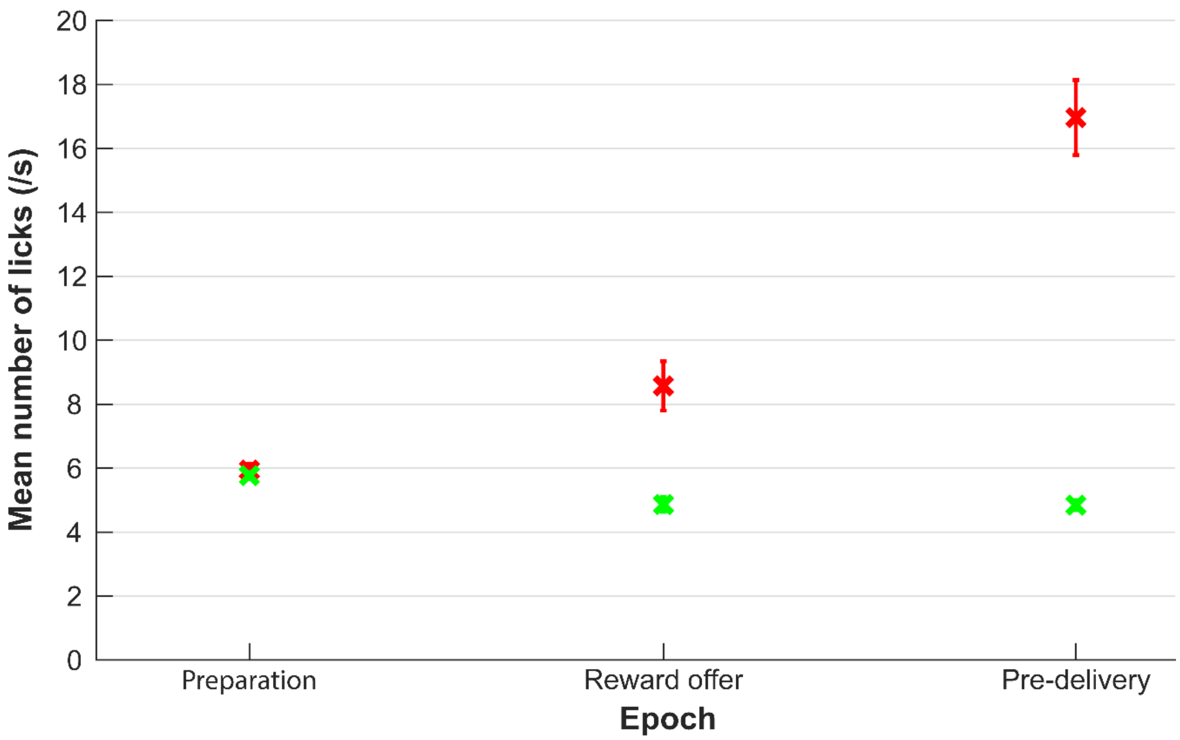


Fig. 4.3) The mean lick frequency is shown for the 0.8ml (red) and 0ml (green) juice rewards for each of the trial epochs in the same PSL session as is shown in fig. 4.2, with 95% confidence intervals for the mean indicated by the bars.

For trials with the 0.8ml predicting fractal, Bonferroni-corrected multiple-comparisons t-tests showed significant differences in mean lick rate between all epochs. The monkey licked more in the ‘Reward offer’ epoch (M = 8.57, SD = 2.72) than in the ‘Preparation’ epoch (M = 5.94, SD = 0.47; $p = 1.23x10^{-4}$), and licked significantly more in the ‘Pre-delivery’ epoch (M = 16.96, SD = 2.97) than in the ‘Reward offer’ epoch ($p = 9.56x10^{-10}$).

On the other hand, Bonferroni-corrected multiple-comparisons tests between mean lick rates in each epoch for the 0ml predicting fractal showed that the monkey licked significantly less in the ‘Reward offer’ (M = 4.86, SD = 0.79) and ‘Pre-delivery’ (M = 4.85, SD = 0.39) epochs than in the ‘Preparation’ epoch (M = 5.76, SD = 0.26; $p = 1.23x10^{-7}$ and $p = 1.57x10^{-6}$ respectively). Interestingly however, there was no significant difference in mean lick rate between ‘Reward offer’ and ‘Pre-delivery’ epochs (p = 0.99), presumably because the animal understood that the red border presentation in the ‘Pre-delivery’ epoch did not indicate the delivery of any reward in the presence of the 0ml predicting fractal.

Sampling of licking responses in a further 3 PSL sessions, in which a third 0.3ml reward and associated fractal were introduced, showed a similar relationship between the volume of reward indicated by the fractal stimulus, the current trial epoch, and the rate of anticipatory licking: In all 3 subsequent sessions, 2-way ANOVA revealed significant effects of both ‘Volume’ and ‘Epoch’ factors, as well as their interaction, on licking rate (results summarised in Table 4.1, with lick data per epoch for each session shown in fig. 4.4).

Table 4.1) Summary of results of a 2-way ANOVA on lick-rate in a set of sessions testing 3 different reward volumes; including effects of trial epoch ‘Epoch’ and reward volume ‘Volume’ factors, as well as their interaction ‘Epoch*Volume’.

Session	Factor	df_{factor}	df_{error}	F-value	P-value
1	Epoch	2	294	114.87	1.37×10^{-37}
	Volume	2	294	67.83	5.99×10^{-25}
	Epoch*Volume	4	294	49.43	8.76×10^{-32}
2	Epoch	2	294	193.68	2.19×10^{-54}
	Volume	2	294	150.37	1.05×10^{-45}
	Epoch*Volume	4	294	45.54	9.39x10-^{30}
3	Epoch	2	294	6.99	1.10×10^{-3}
	Volume	2	294	57.66	7.46×10^{-22}
	Epoch*Volume	4	294	17.35	8.53×10^{-13}

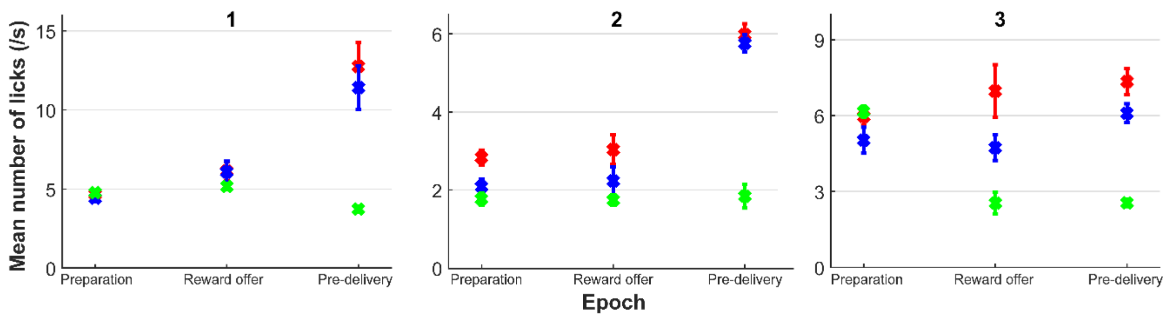


Fig. 4.4) The mean lick frequency is shown for 0.8ml (red), 0.3ml (blue), and 0ml (green) juice rewards for each of the trial epochs in 3 PSL sessions using the same trial structure as is shown in fig. 4.1, with 95% confidence intervals for the mean indicated by the bars.

Although the rate of anticipatory licking was not always significantly higher for the 0.8ml fractal than the 0.3ml fractal, the trend was always in this direction, and the lick-rate for both was always higher than that for the 0ml fractal by the end of the ‘reward offer’ epoch (see fig. 4.4). The lack of reliable differentiation of anticipated reward volume by lick-rate may have been due to the lack of head restraint in these sessions, but the data were enough to show that learning of fractal-reward associations had taken place without requiring any invasive procedures, or training on another choice task, meaning that we could progress to training of the BDM task as quickly as possible.

<u>Pavlovian Stimulus Learning of the water-budget:</u>

Following 12 sessions of training with the fractal stimuli, we moved on to train the monkey with the budget-bar stimulus, whose coloured area indicates the volume of water-budget that the monkey expects to receive on a given trial. Initial training in these sessions used a large volume of budget reward, as did the first version of the preliminary BDM task, however, over the long course of training the BDM and developing and refining the task this budget-bar stimulus would come to represent different volumes of water-budget.

Nevertheless, the same budget stimulus was used throughout BDM testing, and any changes to the identity of the budget were lparnt by the subject over a few sessions of PSL training for the budget stimulus. These sessions always followed the same trial progression that is outlined in figure 4.5, with any timing changes that were made to the BDM task being replicated in these budget-bar training sessions also. Here, we present the task with the timings that were used in preparation for preliminary BDM training.

The budget-bar PSL task progressed similarly to the task used for the training of fractal stimuli. Subjects were first presented with a cross for 0.5s to draw their attention and prepare them for the upcoming trial. They then saw the full budget-bar stimulus, whose total area corresponded to 0.8ml of liquefied food reward, presented on-screen for 1.5s. A green line that would come to represent the computer-bid in the final BDM task was then presented on the budget-bar stimulus, its position being randomly generated from a uniform distribution over the entire range of the budget-bar. After a short delay (0.5s) the area below this green line was occluded with a black rectangle of the same shade as the screen's background, and this area represents the portion of the budget that has been 'spent' in the BDM task. The remaining budget payout, whose volume was proportional to the remaining un-occluded area of the budget-bar was paid out to the subject 0.5s after the occlusion of the bar, with the ITI starting 0.5s after reward receipt. The presence of a white border around the total budget-bar area was intended to allow the monkey to more clearly compare the remaining area (and therefore the anticipated payout) of the budget-bar in the 'Budget offer' epoch with the total area initially presented in the 'Budget-bar' epoch.

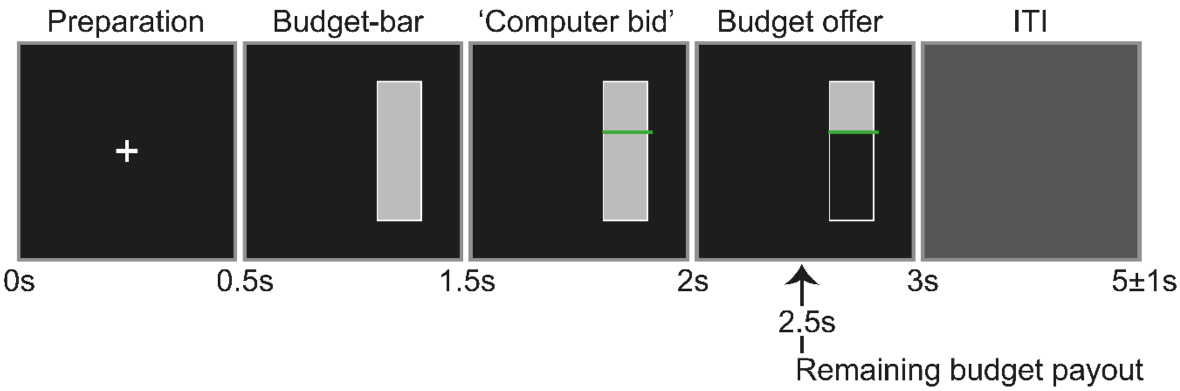


Fig. 4.5) Trials start with a 0.5s preparation period to orient the subject to the screen and the upcoming reward-predicting stimulus. A grey bar with white border (budget-bar) is then shown, indicating the total volume of water-budget. A green horizontal marker that is used to signal the computer's bid in the BDM task is then shown on the bar, and the area below this is soon occluded. A volume of water-budget proportional to the un-occluded area of the budget-bar is paid out 0.5s following the occlusion of the bar in the 'Budget offer' epoch. 0.5s after reward delivery the stimuli disappeared from the screen, and this was followed by a variable inter-trial interval (ITI) of 2±1s, with random durations distributed by a truncated exponential function.

As for PSL training of the fractal stimuli, the interval between the initial presentation of the budget-bar and the receipt of the water-budget was not consistent between this training task and the final BDM task. We reduced this delay as no choice period was necessary during PSL training, and a shorter delay (as well as higher overall reward rates) were more suited to enhancing the speed of learning in the task as well as maintaining motivation over long training sessions. But to maintain the learning of water-budget receipt with respect to relevant task timings, we ensured that the time interval from the appearance of the 'computer-bid' stimulus to the moment of water-budget receipt (1s), and between occlusion of the 'spent' budget below the ‘computer-bid’ and reward receipt (0.5s), was held constant between this training task and the subsequently trained version of the BDM[*].

Following the first two sessions of PSL training for the budget-bar stimulus, we introduced a touch-key (TK) requirement into the task for the remaining 7 sessions[†]. Thus, the monkey was required to hold onto the TK for the duration of the ‘budget-bar’ epoch. TK data was sampled at 50 Hz, with bins of length 0.02s collecting the state of the TK, if the key was being touched then the bin would contain an ‘on’ state indicator, and an ‘off’ state indicator otherwise. The monkey had to maintain hold of the TK for over 90% of the sampled bins for the trial to continue successfully, otherwise, the trial ended without progressing to the ‘Computer bid’ epoch and a blue error screen was shown for 3 seconds instead, followed by the ITI as normal (such that error trials were 1.5s longer than correct trials on average).

Although the monkey did not have to hold the touch-key during the ‘Preparation’ epoch and was free to release it after the ‘Budget bar’ epoch, he only did so on an insignificant number of correct trials, usually holding the TK for the entire duration of correctly performed trials (Table 4.2). On the other hand, the large number of error trials were mostly due to the monkey completely ignoring the TK, and not due to them not holding the TK for long enough – that is, on the vast majority of error trials the TK did not register any touch at all, and observation of the animal during these sessions confirmed that most errors were due to the monkey turning away from the TK and screen completely (the monkeys were not restrained within the primate chair).

Surprisingly, the monkey’s performance with the touch-key saw a sudden huge improvement on the 7th session (Table 4.2), with performance jumping from ~50% in the past two sessions to >90% in that one. After that session, we moved on to the next PSL task, training both fractal and budget-bar together, whilst maintaining the TK requirement.

[*] The BDM task timings did change as we developed the task. Equivalent timing changes were made to all training tasks that were to be used in tandem with the BDM, though once the subjects understood the BDM task minor timing changes could be made without the need for retraining of task stimuli.

[†] Both monkeys were trained to use only their right hand to control the touch-key (and later, joystick) by reaching through a hole in the right-hand side of their experimental primate-chairs.

Unfortunately, we had no way of confirming that learning of the budget-bar stimuli had taken place at this point due to problems with the lick detection hardware, and so we added more sessions of training for the budget-bar stimulus as well as for the combined budget-bar and reward-fractal training (described in the next section) to increase the chances that the contingencies had been learnt. We would only be able to definitively confirm that the monkey had learnt the budget-bar stimulus contingencies at a later stage, using a binary choice task (Ch. 5.1).

Table 4.2) For each session the number of correct trials are shown, as well as the number of trials for which the touch-key (TK) was held throughout all epochs. The number of errors in each session are also shown, as well as the number of trials on which the monkey released the TK too early. On trials in which the monkey held the TK at all, they tended not to release it too early, and often held it throughout the trial – most error trials were due to the monkey completely ignoring the TK. After achieving the high-level of performance seen in session 7, Ulysses moved on to the more complex combined training of budget-bar and reward fractal (fig. 4.6), which still required correct interaction with the touch-key.

Session	Correct	Held throughout	Error	Early TK release	Performance
1	296	292	400	5	42.5%
2	221	217	565	12	28.1%
3	266	260	193	11	58.0%
4	243	241	383	11	38.8%
5	214	207	187	12	53.4%
6	239	235	235	11	50.4%
7	200	199	11	1	94.8%

<u>Combined training of the budget-bar and reward-fractal:</u>

Before moving on to train the monkey to use the joystick for the BDM, we trained him in a final set of 15 PSL sessions in which we presented both the budget-bar and reward-fractal stimuli together, delivering the signalled volumes of juice-reward and water-budget at the end of each trial (fig. 4.6). These sessions were introduced to help the monkey learn the relative timings of reward receipt indicated by the two different stimuli, and, as for the previous PSL sessions, the timing of stimulus changes immediately preceding delivery of the rewards was kept consistent with the timings used in subsequent BDM training. Thus, the remaining budget was still paid out 0.5s after the occlusion of the 'spent' portion of the budget during the 'Budget offer' epoch and the juice-reward was also paid out 0.5s after the appearance of the red border in the 'Pre-reward' epoch.

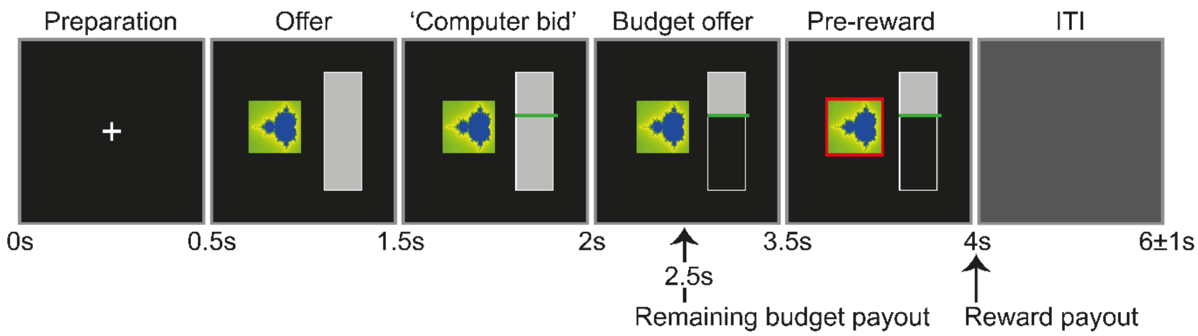


Fig. 4.6) Trials start with a 0.5s preparation period to orient the subject to the screen and the upcoming reward-predicting stimuli. Both the fractal and budget-bar are then shown together for 1s. A randomly generated 'computer bid' then appears and the 'spent' portion of the budget-bar is occluded 0.5s after this. The remaining budget is paid out a further 0.5s after the occlusion of this 'spent' portion. 1 second after the budget payout a red border appears around the fractal and the signalled reward is paid out 0.5s later. Reward payout was followed by a variable inter-trial interval (ITI) of 2±1s, with random durations distributed by a truncated exponential function.

Comparing this to previous training steps, though the time between the first presentation of the fractal stimulus and the delivery of the reward has been significantly increased (1.5s in the PSL task for the reward stimulus alone vs. 3.5s in this combined task), the time between appearance of the border and delivery of the reward is held constant.

As for training of the budget-bar stimulus, we also required that the monkey maintain hold of the touch-key for the duration of the 'Offer' epoch (equivalent to 'Budget-bar' epoch in fig. 4.5), and by this point in the monkey's training his ability to do this reliably had vastly improved, with correct performance of the touch-key requirement in over 90% of trials in all 15 of the combined PSL sessions.

Operant conditioning of joystick control – target task:

The final training step before introducing the preliminary BDM procedure was an operant conditioning task designed to teach the monkey the visuomotor association between the forward/backward movement of a custom-made joystick and the up/down movement of an on-screen marker. This would come to form the basis of bidding in the BDM, with the monkey's bid being determined by the final position of a red 'bid-marker' on the budget-bar. We therefore wanted the monkey's movements of the joystick and bid-marker to be as well trained and controlled as possible, as their degree of motor precision would directly impact the degree of certainty with which a value could be inferred from a given distribution of their bids. Moreover, acquiring a profile of the monkey's motor variability could allow us to better appreciate the limits of bidding precision/consistency that are imposed by this source of noise. We achieved this by training the monkey to move this bid-marker into a target region on the budget-bar in a specially designed 'target task'.

On a given trial (fig. 4.7), following a short 'Preparation' period, the monkey's bid-marker appeared at the centre of the budget-bar in a dark shade of red, as did a randomly located blue rectangular target region. The bid-marker then changed colour to a brighter shade of red at the beginning of the 'Movement' epoch, indicating that its position could now be manipulated by the joystick – the monkey then had 2s to move the bid-marker into the target region, and was rewarded on trials in which they achieved this – receiving no rewards at all in case of a miss.

Trial progression in the target task was designed to mimic BDM trials as closely as possible. For example, the epochs in the target task were all the same length as the equivalent epochs in the subsequent BDM task[*] (fig. 4.10), and the timing of reward and budget receipt were consistent relative to the stimulus changes that were common to both tasks, e.g. delivery of juice-reward 0.5s after appearance of the red border in the 'Pre-reward' epoch of both tasks.

Importantly, this task also constituted the monkey's first introduction to the win/loss contingencies of the BDM. The final position of the bid-marker was taken as a BDM bid whenever it landed within the target region. Following this, a randomly generated

[*] Although these timings would later be changed as the BDM task was developed.

green computer-bid marker appeared and the reward to be delivered was determined as in the BDM: If the computer-bid was lower than the monkey's bid-marker, then the budget-bar below the 'computer-bid' marker was occluded and the monkey received both the remaining budget, represented by the remaining un-occluded portion of the budget-bar, as well as the reward signified by the fractal stimulus (fig. 4.7a). And if the computer-bid was higher than the monkey's bid-marker, then the whole budget-bar remained un-occluded and the monkey received the full amount of water-budget, but not the reward indicated by the fractal (fig. 4.7b).

In this early version of the target-task, as well as in the preliminary BDM task, the behavioural requirements for trial progression were kept minimal as the demands of correct target placement were already high (and this was reflected in the large proportion of trials on which the monkey made touch-key errors – see Table 4.3). As such, the only behavioural requirement beyond correct marker placement was that the monkey had to maintain hold of the joystick* during the 'Offer' epoch, consistent with previous PSL training sessions. If the joystick was not held throughout this epoch, then a blue error screen was shown for 6s at the end of the epoch followed by the grey ITI screen before the next trial.

* The joystick was adapted by "Biotronix" to have an incorporated touch-key function.

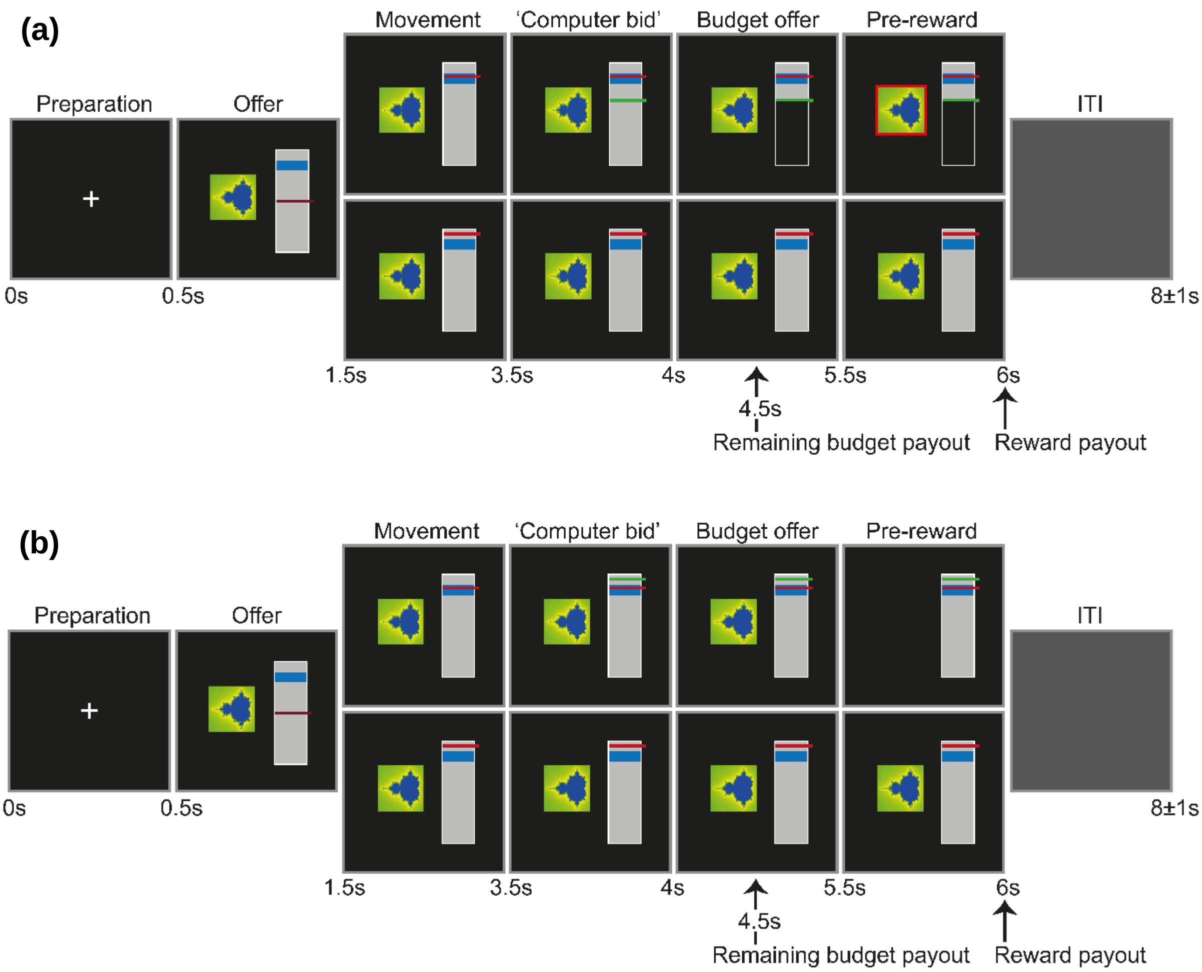


Fig. 4.7) a) Trials started with a 0.5s preparation period to orient the subject to the screen. Then, during the 'Offer' epoch, both the fractal and budget-bar are shown for 1s alongside the starting position of the monkey's bid-marker (dark red) and the blue target region. At the start of the 'Movement' epoch the bid-marker changes shade to a bright red to signal that movements of the joystick may be made: The monkey then has 2s to move the bid-marker into the target region. The position of the marker at the end of the 'Movement' epoch determines whether the trial was a hit (bid-marker inside target region – top), or a miss (bid-marker outside target region - bottom). In the case of a successful target hit, a randomly generated computer-bid marker appeared and the reward outcomes and subsequent stimulus changes were determined as in the previous 'combined PSL training' and in the subsequent BDM task. In the case of a miss the monkey had to wait for the same duration – ensuring a reduced overall reward-rate - but no further stimulus changes occurred and no reward or budget were paid out. The final 'Pre-reward' epoch was followed by a variable inter-trial interval (ITI) of 2±1s, with random durations distributed by a truncated exponential function. **b)** As for (a) but showing trial progression when the 'computer-bid' is higher than the monkey's 'bidding' marker. Note the disappearance of the fractal stimulus at the equivalent time to appearance of the red border at the start of the 'Pre-reward' epoch' in (a).

Initially, we trained the monkey on a version of this task with targets of height equal to 1/3rd of the height of the budget-bar and progressively reduced the target size until it was equal to only 1/20th of the budget-bar height. In this phase of training, our focus was on achieving precise movements of the bid-marker as quickly as possible, so we often reduced the target size several times in a single session if it was immediately clear that the monkey was performing well at a given target size.

However, to better characterise the monkey's motor variability, as well as any improvements in their performance across sessions (fig. 4.8, Table 4.3), we went on to collect a data-set of 8 sessions consistently using the smallest target size of 1/20th of the budget-bar height and excluding trials in which the monkey made a touch-key (TK) error, i.e. where the joystick was not held during the 'Offer' epoch, focusing only on those trials in which an attempt to move the marker was made. Of these attempted trials, the proportion in which the monkey correctly placed the marker in the target gave an 'Attempted rate', and we reached an overall level of ~90% performance in attempted trials before moving on to the preliminary BDM task.

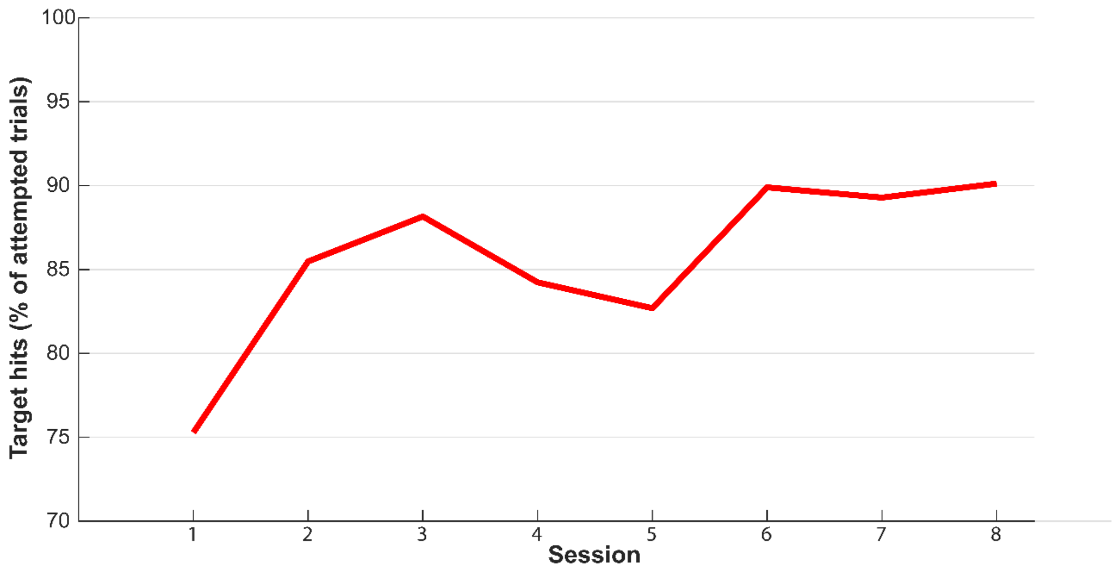


Fig. 4.8) Showing the percentage of trials in which the monkey successfully placed the bid-marker within the blue target region after excluding trials in which a touch-key error was made.

Table 4.3) Summary of performance in the final set of sessions shown in fig. 4.8, and for which a consistent target size of 1/20th of the budget bar height was used. 'Attempted' trials are those on which no touch-key (TK) error occurred and the monkey tried to move the bid-marker with the joystick. The 'Attempted rate' is the proportion of attempted trials on which the bid-marker was correctly placed in the target region, and 'Overall performance' is simply the proportion of all trials on which this occurred. Number of trials for each category are shown in brackets.

Session	Attempted rate	TK error rate	Overall performance
1	75.3% (201)	4.0% (11)	72.3%
2	85.5% (384)	4.9% (19)	85.5%
3	88.2% (335)	14.6% (65)	75.3%
4	84.2% (358)	15.3% (77)	71.3%
5	82.7% (172)	72.6% (552)	22.6%
6	89.9% (338)	0.5% (2)	89.4%
7	89.3% (350)	5.5% (23)	84.3%
8	90.1% (301)	11.4% (43)	79.8%

A clear improvement in performance on attempted trials is observed between the first and last sessions using the smallest target size, and this is most pronounced for targets that are away from the extremes of the budget-bar (fig. 4.9). The improvement in performance could also be measured in terms of the mean absolute deviation of bid-marker placement from the target centre, and a 1-way ANOVA found a significant effect of session on this [$F(7,2739) = 5.45$, $p = 3.13 \times 10^{-6}$], with a Bonferroni-corrected multiple-comparisons test showing a significantly reduced mean absolute deviation from target centre by session 6 ($M = 0.011$, $SD = 0.012$) compared to session 1 ($M = 0.02$, $SD = 0.04$, $p = 7.6 \times 10^{-3}$), though mean absolute deviation was not significantly less in session 8 ($M = 0.013$, $SD = 0.023$, $p = 0.16$).

However, due to *bid-censorship* effects we wanted to avoid bids at either extreme - particularly at the top of the budget-bar - and were therefore most concerned with improvements in performance on more centrally located targets. Excluding bid-marker placements for these targets, using Bonferroni-corrected multiple-comparisons t-tests, the mean absolute deviation from target centre is now found to be significantly lower in session 6 (M = 0.010, SD = 0.010), session 7 (M = 0.010, SD = 0.011), and session 8 (M = 0.010, SD = 0.015) when compared to session 1 (M = 0.019, SD = 0.026, all p<0.02).

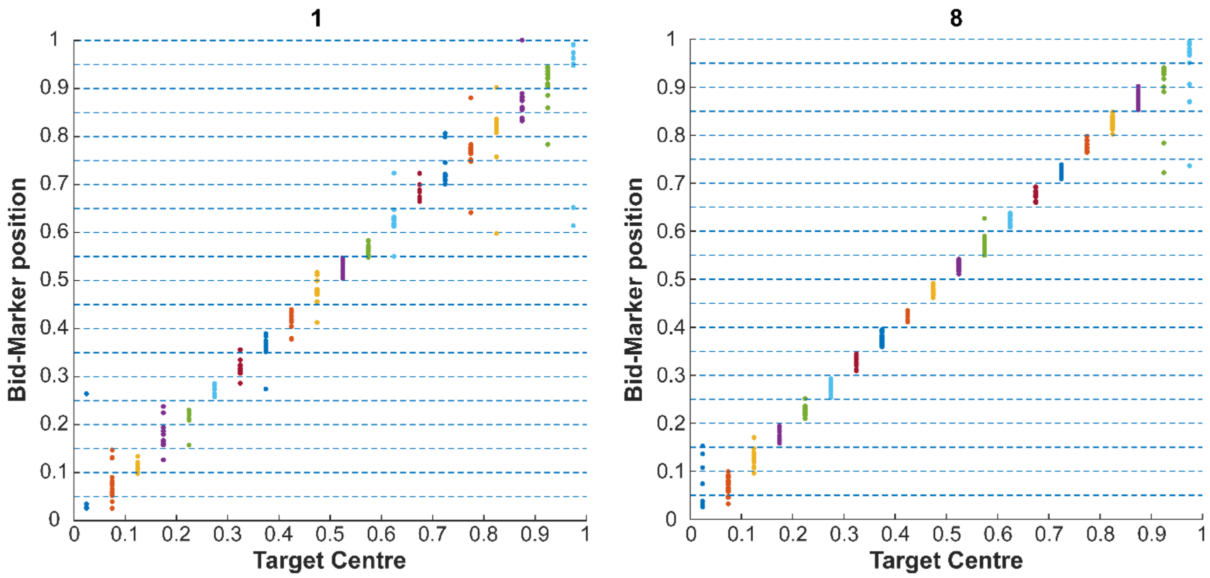


Fig. 4.9) Bid-Marker positions for each target are shown for two sessions from the final target-task training set; the first and last sessions. The boundaries of the target region relative to the bid-marker position are shown by the blue dashed lines, and both target and bid-marker positions are given as relative to the total bar-height (e.g. a bid-marker at 0.25 is positioned a quarter of the way up the budget-bar). The degree of deviation from the centre of the target is clearly reduced by the 8th session, though less so at the extremes of the budget-bar.

By session 8 of the target task using the smallest target size, the monkey's performance had reached an asymptotic level at ~90% on attempted trials. The early version of the target-task described here was therefore sufficient for teaching the monkey how to control an on-screen marker with a joystick, as well as exposing the monkey to the contingencies of the BDM. And whilst the few behavioural requirements outside of correct bid-marker placement were useful in allowing the monkey to more easily acquire the visuomotor association, it also meant that movements were poorly controlled.

Once the monkey had learnt to control the joystick, it became possible to introduce further, more rigorous, behavioural requirements as we sought to improve the quality of bids by more tightly controlling the monkey's movement. These additional requirements included non-movement of the joystick in the 'Offer' epoch and the need to stabilise the bid-marker for 0.25s in the 'Movement' epoch, bringing it to rest at its final position to record a bid or a target-hit and these changes were applied both to later versions of the target-task, as well later versions of the BDM (Ch. 5.1).

However, at this stage we prioritised establishing whether the BDM was at all viable, and therefore moved on to the preliminary BDM task using the same minimal behavioural requirements as described for the early version of the target-task described here.

2 – Preliminary BDM testing

The initial training stages described in the preceding section taught the monkey how to associate specific stimuli with the delivery of specific magnitudes of reward at certain times; how to move the bid-marker to an intended location using a joystick; and, introduced them to the contingencies of reward delivery given the position of their bid-marker relative to that of the computer's. Now that these basic informational components of the BDM had been acquired, we started training the monkey in an early version of the task.

While many aspects of the task would come to change between this preliminary version and the final refined version of the BDM (including epoch timings, stimuli, behavioural requirements of joystick control etc. - Ch. 2.2), the essential task structure (fig. 4.10) stayed the same.

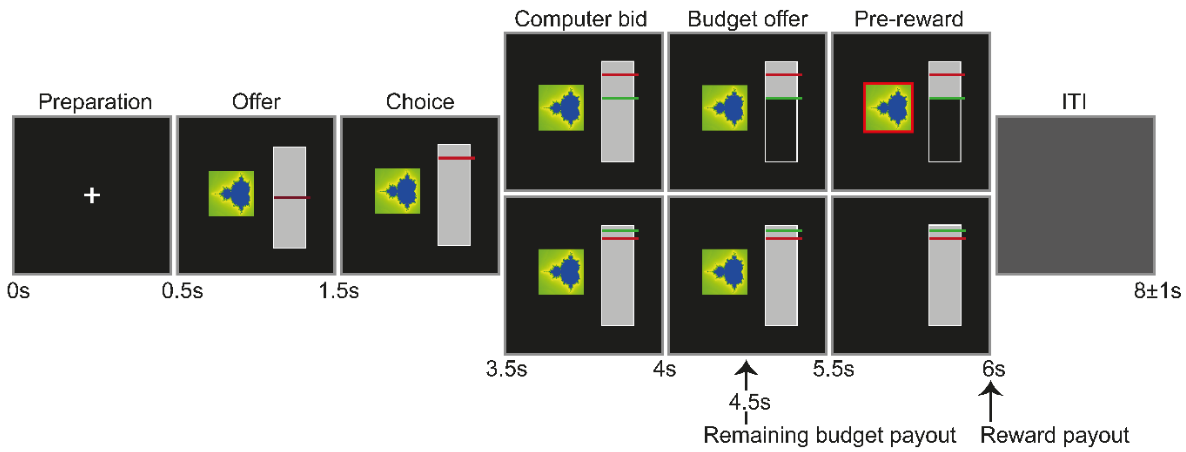


Fig. 4.10) Trial structure of the preliminary BDM task. Task timings and contingencies were consistent with those used in the preceding target-task (see figure 4.7). The 'Choice' epoch replaces the 'Movement' epoch of the target-task – it is of the same duration, but instead of having to move the bid-marker to a specified target region, the monkey is free to move it to any location.

The key tension in the task design at this stage was in the choice of the volume to be used for the budget. On the one hand, a high budget volume means that a given change in the position of the bid-marker on the budget-bar will lead to a bigger change in absolute payoff when compared to a lower budget volume. Making the absolute differences in payoff for different bids larger should make those differences more perceptible and encourage more careful bidding. This is particularly important when one considers the relatively low costs of deviating from optimal behaviour in the BDM (Appendix 1).

On the other hand, a lower budget volume means that the subject can complete more trials in each session before reaching satiety, which could accelerate learning of the task by allowing the monkey to experience the task contingencies and payoffs on more occasions. Moreover, increasing the number of trials to satiety would have practical benefits for any follow-up neurophysiological studies (Ch. 6.1), as each neuron that was isolated for recording could be tested more extensively.

We collected data from versions of the task using both high (1.5ml) and low (0.8ml) budget volumes during these preliminary stages of training, and present those results in the sub-sections that follow.

Two other factors of the task design were also manipulated during these preliminary training sessions: the distributions from which the computer drew its bid, and the starting position of the bid-marker.

Our first two versions of the task only tested different budget volumes and used a uniform distribution for the computer's bids, but a later variant also tested the feasibility of using different 'peaked' computer-bid distributions to manipulate the expected costs of misbehaviour (ECM), and, importantly, lent support to the idea that the monkey's bids would not come to be simply determined by the concentration of computer bids to a certain range of bids- for example, through some heuristic, such as matching of the bid-marker position to the mean computer bid-marker position.

The final change that came about as a result of this preliminary BDM testing was to the starting position of the bid-marker. Whilst we had speculated that a central starting position would be best (as shown in figure 4.10) so as not to anchor or bias the monkey's bids towards one end of the bidding range or the other, we found that this led to a huge proportion of bids being placed at the very top of the bidding-range, with very few downwards movements of the bid-marker being made. This suggested a preference for forward movement of the joystick (translating to upward movement of the bid-marker). We therefore changed the bid-marker starting position to be at the bottom of the budget-bar in the 'Offer' epoch, such that the monkey could make better use of the whole range of possible bids*.

The following 3 sub-sections describe the results of our initial testing of the first monkey, Ulysses, using these different task parameters.

Higher budget volume task:

We used a Chi-squared test to ensure that the computer-bid distribution that the monkey faced in this condition was in fact uniform for each of the two reward magnitudes used in these sessions. Pooling computer bids from all 10 sessions in this data-set, we were unable to reject the null hypothesis that the computer bid distributions were uniform for the high magnitude reward, $\chi^2 = 17.0$, $p = 0.98$, or for the low magnitude reward, $\chi^2 = 17.4$, $p = 0.99$. Moreover, a two-sample t-test showed that there was no significant difference between the mean computer bids for the high volume (M = 0.78, SD = 0.43) and the low volume rewards (M = 0.79, SD = 0.42, $p = 0.75$). This was also confirmed on a session-by-session basis, where two-sample t-tests found a significant difference between computer bids for the two juice rewards on only one of the 10 sessions (session 8).

Looking at trials from across all sessions we also found no significant difference between the mean subject bids for the high (M = 1.20, SD = 0.30) and low magnitude rewards (M = 1.21, SD = 0.29, $p = 0.52$). In agreement with this, a Spearman's rank correlation found no significant effect of reward magnitude on the monkey's bids (Rho = - 0.011, $p = 0.74$).

However, an analysis of the speed of bid-marker movement over the choice period, measured in pixels/second, shows a significantly greater speed of bid-marker

* Ultimately, we would like to use a random bid-marker starting position, as in the final version of the BDM task (Ch. 2.2), especially as this would preclude the monkey preplanning his motor response; something that is especially important for neuronal recording. However, at this stage we felt that the demands of such a behavioural requirement would be too great, and first sought to improve the monkey's bidding using a simpler paradigm.

movement for the high magnitude reward (M = 195.4, SD = 141.7) than for the low magnitude reward (M = 174.5, SD = 126.7, $p = 0.015$). This is indicative of the fact that despite there being no significant difference between bids for either reward, the monkey was still discriminating between the two reward magnitudes with the vigour of his motor response, as has been described elsewhere[71,72].

An analysis of individual session data shows that there was a significant difference between bids for the high and low magnitude rewards on 4 of the 10 sessions; sessions 1, 4, 7, and 10 (fig. 4.11a). However, this was in the opposite direction to that expected for the first 3 of these sessions, with higher bids for the higher magnitude reward only being observed in session 10 (High magnitude reward: M = 1.47, SD = 0.18; Low magnitude reward: M = 1.36, SD = 0.29, $p = 0.008$).

Closer inspection of the monkey's bids shows that this difference was mainly driven by non-movement of the bid-marker on a subset of sessions for the low-magnitude reward (bids at 0.75 are at the central starting location of the bid-marker), with almost all the other bids being placed at the top of the budget-range; effectively making the behavioural response a binary one (see session 10 in fig. 4.11b).

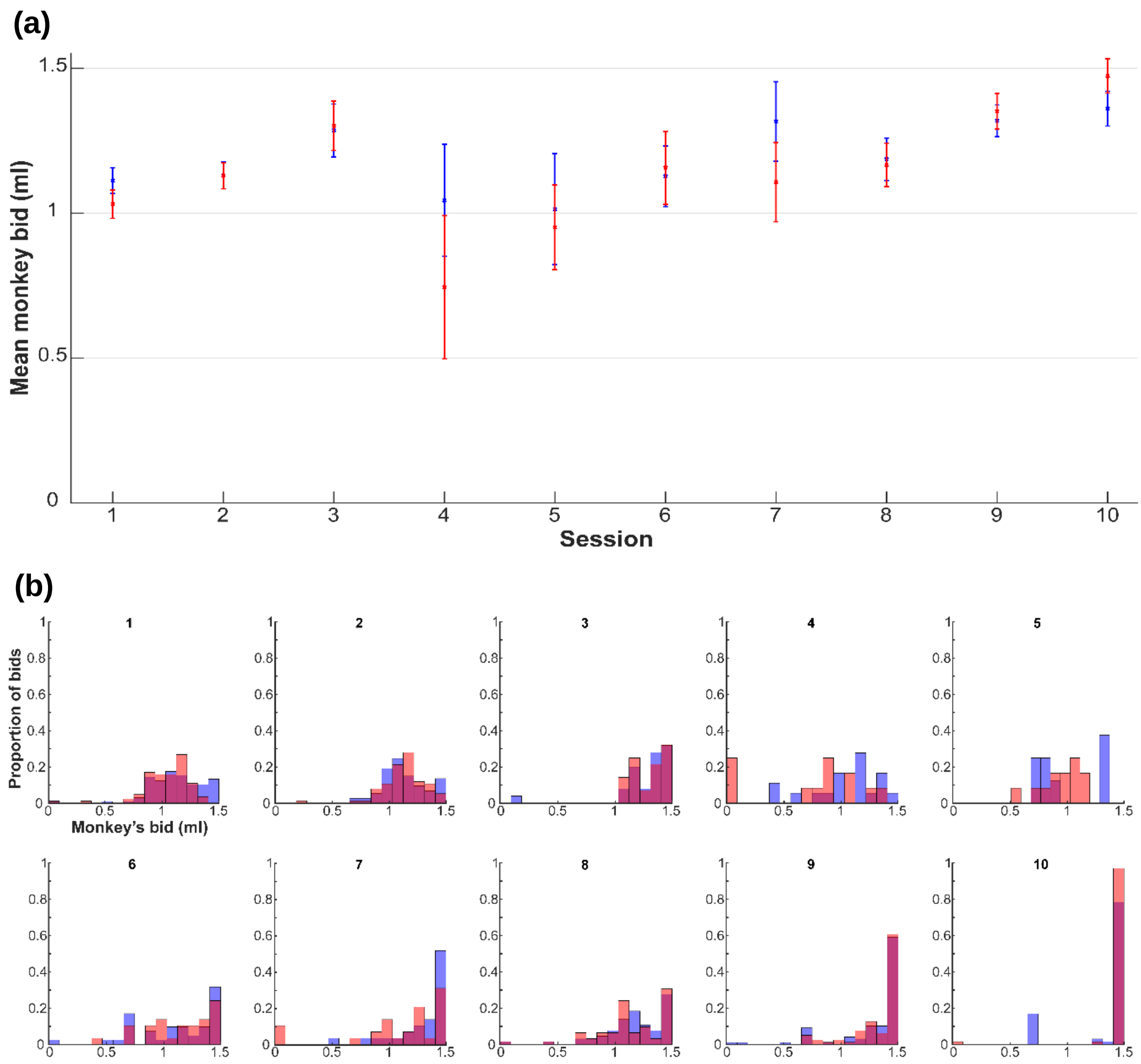


Fig. 4.11) a) Mean bids for high (0.4ml, red) and low (0.1ml, blue) reward magnitudes over 10 sessions of the preliminary BDM task using a budget volume of 1.5ml, [1.5-BDM]. Error-bars are 95% confidence intervals for the mean. **b)** A probability histogram of the data for each session shown in (a); the monkey showed a propensity for bidding at the very top of the range for both the high (0.4ml, red) and low (0.1ml, blue) reward magnitudes.

Table 4.4) Summary data for the 10 sessions shown in figure 4.11, giving means (M, in ml), standard deviations (SD) and number of trials (n) for both high magnitude (H) and low magnitude (L) rewards for each session. Performance shows the percentage of trials that were correctly completed, i.e. where the touch-key was held for the duration of the 'Offer' epoch.

Session	H - M	L - M	H - SD	L - SD	H - n	L - n	Performance
1	1.03	1.11	0.21	0.24	82	97	86.9%
2	1.13	1.13	0.20	0.19	75	73	75.9%
3	1.30	1.29	0.16	0.28	28	25	26.5%
4	0.74	1.04	0.48	0.31	12	18	13.0%
5	0.95	1.01	0.19	0.28	12	8	19.6%
6	1.16	1.13	0.29	0.35	29	41	32.4%
7	1.11	1.32	0.44	0.25	29	29	32.6%
8	1.17	1.19	0.31	0.28	62	65	56.4%
9	1.35	1.32	0.23	0.31	79	98	69.7%
10	1.47	1.36	0.18	0.29	67	65	70.6%

There was a prevalence of bidding at the very top of the budget-bar range for both reward magnitudes (figure 4.11b), and we speculated that this may have been driven by a simple strategy of placing maximum bids. Such a strategy would place minimal cognitive demands on the monkey and would guarantee acquisition of the juice-reward on every trial, and, a sufficiently high reward rate such that the monkey was often sated within 200 trials. We reasoned that if such a strategy was motivated by an avoidance of the cognitive costs associated with accurately placing a bid, then it should become more prevalent as the monkey's overall motivation in the session diminished. To investigate this possibility, we performed a Spearman's rank correlation between the proportion of bids at the top of the budget-bar range (defined as those bids made in the top 1% of the budget-bar) and the cumulative volume of juice that had been consumed - assuming that the degree of motivation to seek reward/the reward's utility is inversely related to satiety[73,74,75,76,77].

We found a significant positive relationship between cumulative consumption of juice and the proportion of bids made at the top of the budget-bar at the population-level (Rho = 0.24, $p = 1.13 \times 10^{-14}$). Looking at individual sessions, a significant positive relationship was also found in 7 of the 10 sessions, no relationship was found in 2 of the sessions (for session 5 no bids were made in the top 1% of the budget-bar range*), and a significant negative relationship was found in session 3 (Fig 4.12, Table 4.5). Importantly, the last 5 sessions all showed a strong significant positive relationship, and it is possible that this was a strategy that the monkey had developed over time such that the relationship only became more clearly apparent in later sessions.

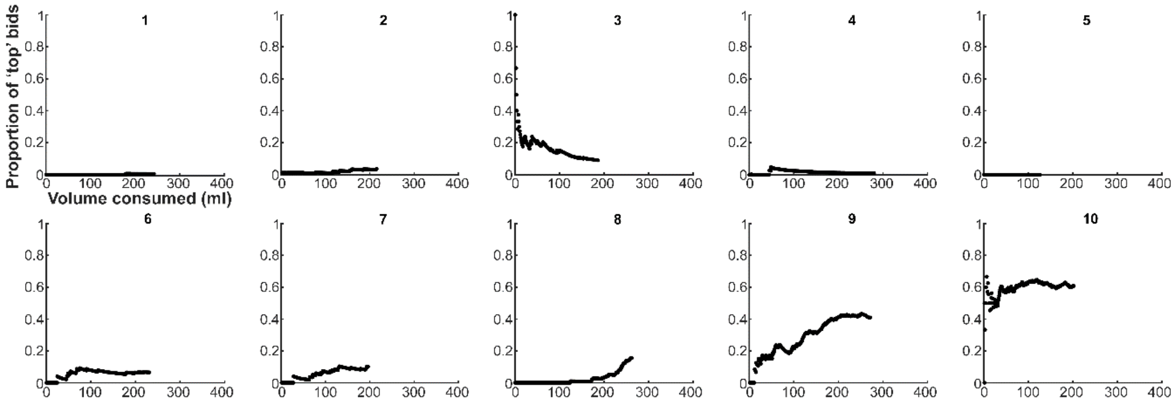


Fig. 4.12) For 7 of the [1.5-BDM] sessions, a significant positive relationship exists between the proportion of bids that the monkey placed in the top 1% of the bidding range and the total volume of reward that the monkey had consumed in that session, with an increasing propensity to place bids at the top of the range in later sessions.

* Sessions 4 and 5 both had fewer than 50 correct trials each, but have been included for completeness and continuity in terms of training progression.

Table 4.5) Results of a Spearman's rank correlation between the proportion of bids placed in the top 1% of the bidding range and the total volume of consumed reward for the data shown in figure 4.12.

Session	Spearman's Rho	p-Value
1	0.73	9.82×10^{-31}
2	0.94	1.38×10^{-71}
3	-0.88	8.73×10^{-18}
4	-0.16	0.41
5	N/A	N/A
6	0.38	1.1×10^{-3}
7	0.95	2.46×10^{-30}
8	0.96	3.30×10^{-72}
9	0.98	1.31×10^{-121}
10	0.47	1.47×10^{-8}

The volume of water-budget used in the task is inversely related to the number of trials for which the monkey works before becoming sated, meaning that with higher water-budget volumes the monkey will experience fewer auctions and outcomes each day, and would therefore be expected to learn to bid optimally at a slower rate, especially considering the low costs of deviating from optimality in the BDM. Given this theoretical concern and the evidence of such satiety effects on the monkey's bidding strategy, we proceeded by reducing the budget volume from 1.5ml of water to 0.8ml of water reward – with the goal of increasing the number of trials in a session before satiety would affect the monkey's behaviour and motivation.

We continued with this new lower budget volume, [0.8-BDM], for a further 13 sessions. The results from that set of sessions are shown in the next section.

Lower budget volume task:

As for the [1.5-BDM] task, we first confirmed that the computer bid distribution that the monkey faced in this condition was uniform for each of the two magnitudes used. Using a Chi-squared test for uniformity, and pooling computer bids from all 13 sessions for this condition, we found no significant difference between the uniform distribution and that of the computer bids for either the high magnitude reward, $X^2 = 13.9$, $p = 0.47$, or the low magnitude reward, $X^2 = 10.5$, $p = 0.22$. At this population level, there was also no significant difference between mean computer bids for the high volume ($M = 0.40$, $SD = 0.23$) and low volume rewards ($M = 0.40$, $SD = 0.23$, $p = 0.47$). This was further confirmed in individual sessions, where a Bonferroni-corrected multiple-comparisons test at the standard alpha level of 0.05 found no significant difference between computer bids for the two juice rewards in any of the 13 [0.8-BDM] sessions.

At this point, our analysis of these sessions departed from that of the preceding [1.5-BDM] task, as we did find a significant difference between the mean subject bids for the high ($M = 0.76$, $SD = 0.11$) and low magnitude rewards ($M = 0.69$, $SD = 0.20$, $p = 3.84 \times 10^{-26}$) across sessions. And a Spearman's rank correlation also showed a weakly positive but significant effect of reward magnitude on the monkey's bids ($Rho = 0.15$, $p = 2.68 \times 10^{-15}$). Alongside this, an analysis of the speed of bid-marker movement over the choice period, measured in pixels/second, continued to show a significantly greater speed of bid-marker movement for the high magnitude reward ($M = 306.3$, $SD = 136.3$) than for the low magnitude reward ($M = 266.5$, $SD = 148.8$, $p = 2.35 \times 10^{-13}$).

Critically, conducting two-sample t-tests on individual sessions showed significantly higher mean bids for the high magnitude reward in 10 of the 13 sessions, with no significant differences in sessions 1, 3, and 7 (fig. 4.13) - and these results hold without assuming equal variances (Appendix 3). As for the population level, the Spearman's rank correlation also showed a weakly positive relationship between the reward magnitude and the monkey's bid which was significant in 9 of the sessions, and was non-significant in sessions 1, 3, 4 and 7 (the results of this analysis are summarised in Table 4.6). However, as for the earlier training sessions with the [1.5-BDM], the difference in bids for the two volumes of juice reward seems to have been largely driven by a difference in the proportion of bids made to the very top of the bidding range, with little difference in bids for the two magnitudes otherwise (fig 4.13b) – and in support of this, exclusion of bids placed in the top 1% of the bidding range prevents differentiation of bids on the basis of reward magnitude by a two-sample t-test in sessions 5, 11, 12, and 13.

However, unlike in the [1.5-BDM] sessions, at the population level a Spearman's rank correlation showed a significant negative (rather than positive) relationship between the proportion of bids placed in the top 1% of the bidding range and the cumulative volume of juice that had been consumed (Rho = -0.12, $p = 5.27 \times 10^{-11}$), and we now found no consistent relationship between these two factors when analysing individual session data: of the 13 sessions, 5 showed a significant positive relationship, 1 showed no correlation between these two factors (session 3), and the remaining 7 sessions showed a significant negative relationship. Moreover, there was no pattern to these different relationships across sessions (e.g. a more consistent positive relationship developing in later training sessions, as in figure 4.12).

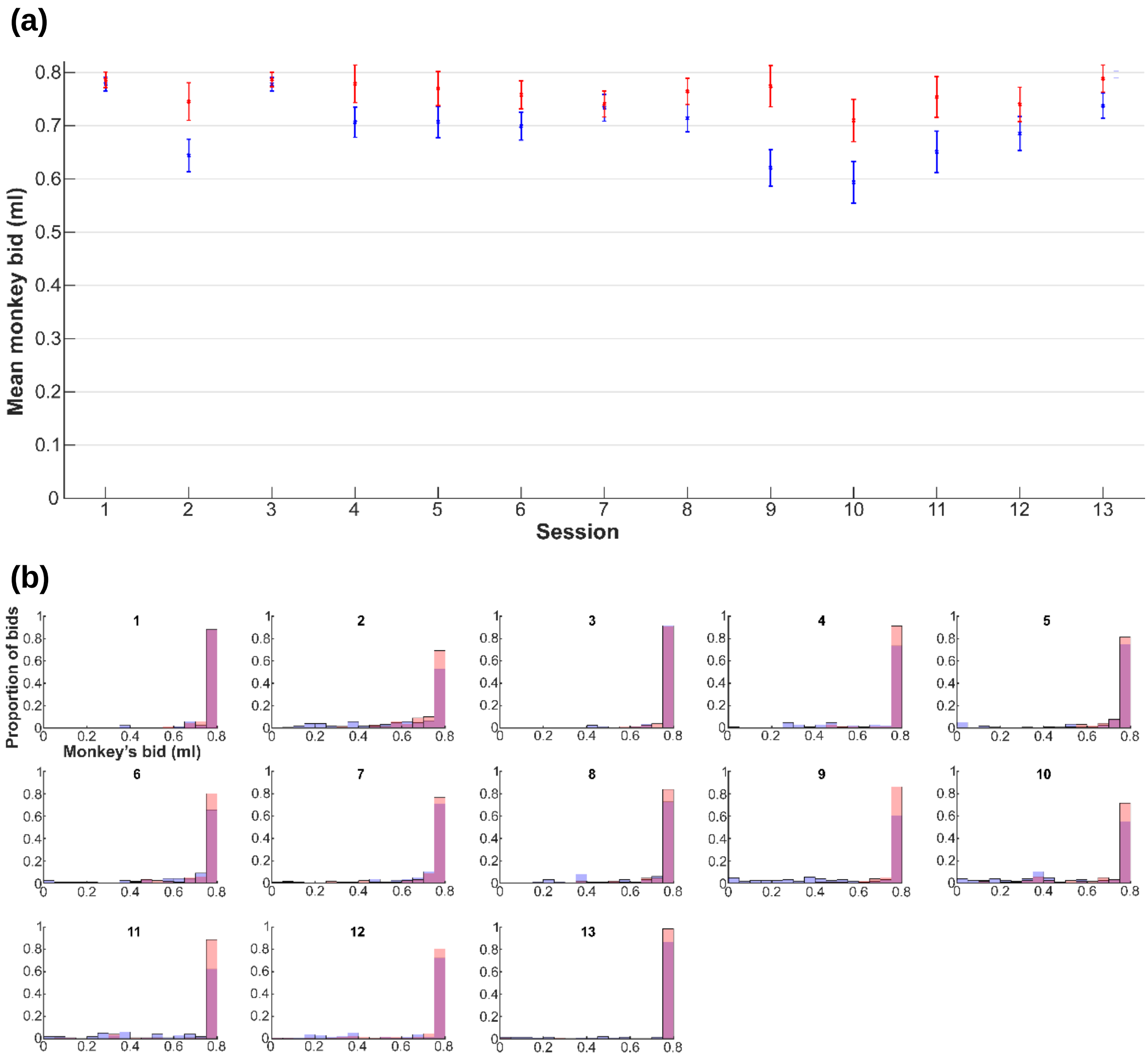


Fig. 4.13) a) Mean bids for high (0.4ml, red) and low (0.1ml, blue) reward magnitudes over 13 sessions of the preliminary BDM task now using a lower budget volume of 0.8ml, [0.8-BDM]. Error-bars are 95% confidence intervals for the mean. **b)** A probability histogram of the data for each session shown in (a); the monkey persisted in showing a propensity for bidding at the very top of the range for both the high (0.4ml, red) and low (0.1ml, blue) reward magnitudes.

Table 4.6) Summary data for the 13 sessions shown in figure 4.13, giving means (M, in ml), standard deviations (SD) and number of trials (n) for both high magnitude (H) and low magnitude (L) rewards for each session. Performance shows the percentage of trials that were correctly completed, i.e. where the touch-key was held for the duration of the ‘Offer’ epoch. Also shown are the results of a Spearman’s rank correlation (Rho and p-value) between the monkey’s bid and the reward magnitude.

Session	H - M	L - M	H - SD	L - SD	H - n	L - n	Rho	p-value	Performance
1	0.79	0.78	0.039	0.073	67	84	-0.058	0.94	66.8%
2	0.75	0.64	0.091	0.21	93	123	0.17	0.014	71.5%
3	0.79	0.78	0.04	0.073	77	82	0.018	0.83	69.7%
4	0.78	0.71	0.069	0.18	68	102	0.13	0.094	55.9%
5	0.77	0.71	0.068	0.21	96	116	0.20	0.0044	68.4%
6	0.76	0.70	0.086	0.19	121	122	0.15	0.021	73.4%
7	0.74	0.73	0.14	0.14	124	117	0.053	0.41	68.9%
8	0.76	0.71	0.089	0.16	111	99	0.15	0.036	70%
9	0.77	0.62	0.065	0.26	101	129	0.28	2.2×10^{-5}	76.7%
10	0.71	0.59	0.17	0.27	123	127	0.20	0.0012	83.3%
11	0.75	0.65	0.15	0.23	97	96	0.35	6.8×10^{-7}	63.5%
12	0.74	0.69	0.15	0.21	129	130	0.13	0.037	86.3%
13	0.79	0.74	0.087	0.18	123	142	0.20	0.0012	85.5%

While these sessions did show an improvement in the ability to differentiate rewards by the monkey's bidding behaviour, the bids remained inconsistent, showing only a weak relationship between the reward magnitude and the monkey's bids. Moreover, the high proportion of bids made to the top of the budget-bar range meant that a high degree of bid censorship made the overall distribution of bids less informative.

Nevertheless, this session set did suggest for the first time that the BDM method developed here could be used to determine a monkey's reward preferences, though improvements were needed to encourage more meaningful bids. Though it was unclear whether it was the reduction in budget volume or greater experience of the monkey that led to the improvements in bidding for this most recent data set (note that for the last session of the [1.5-BDM], the monkey's bids were significantly greater for the higher reward magnitude), we chose to maintain the current budget volume of 0.8ml of water given the aforementioned theoretical and practical concerns. Perhaps most importantly, a two-sample t-test showed that the monkey completed a significantly greater number of trials per session with the lower 0.8ml budget volume ($M = 215.3$, $SD = 37.8$) than when the higher 1.5ml budget volume was used ($M = 99.4$, $SD = 60$, $p = 1.24 \times 10^{-5}$).

We now sought to improve the monkey's bidding by utilising computer-bid distributions that would further increase the costs of sub-optimal bidding behaviour, otherwise using the same task parameters as in this [0.8-BDM] version of the task.

<u>Testing 'peaked' computer-bid distributions:</u>

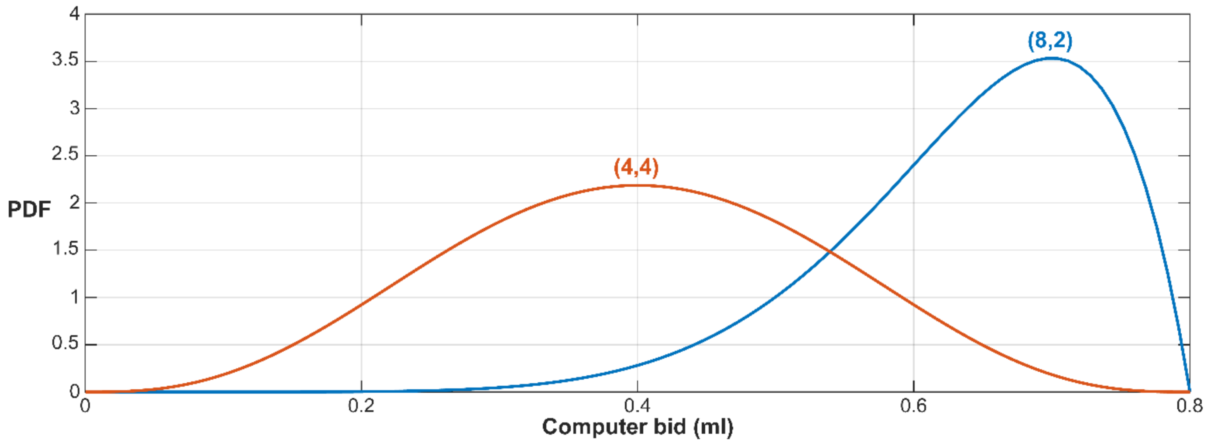
The task structure used for testing non-uniform, 'peaked', computer-bid distributions was the same as for the two preceding session-sets, and as outlined in figure 4.10. Given the monkey's ability to use their bids to successfully distinguish two different reward magnitudes in the immediately preceding task, we also introduced two more reward fractals for this set of sessions: a no-reward predicting fractal, which the monkey had seen before in Pavlovian training sessions, and a new fractal predicting a 0.2ml reward of the same blackcurrant juice as used before (as for changes to the budget volume, the monkey underwent several sessions of PSL training with the new fractals alone and in combination with the budget-bar before their incorporation into the BDM task). Thus, we gradually increased the complexity of the task, but the key difference for these sessions was that the computer drew bids from pre-determined non-uniform beta distributions.

Ultimately, we intended to use such computer-bid distributions to increase the expected costs of misbehaviour (ECM) and thus accelerate the learning of optimal bidding behaviour (Ch. 2.1, summarised in Appendix 1), but we first had to ensure that the monkey was not simply matching their bids to the computer-bid mean – as this would provide a feasible mechanism by which the monkey would appear to differentiate the different reward magnitudes with their bids, but without learning, or responding to, the task contingencies themselves.

Therefore, we first tested a set of 6 sessions using a 'grouped peaks' condition, [4R-BDMp], whereby the 0.4ml and 0.2ml rewards (group 1) had their computer bids drawn from the same (8,2) Beta distribution and the 0.1ml and 0ml rewards (group 2) had their computer bids drawn from another (4,4) Beta distribution (fig. 4.14a). Critically, in this session-set, we were investigating whether the monkey's bids would still differ amongst rewards in the same computer-bid group, i.e. those that drew their computer bids from the same distribution.

First, we checked that the realised distributions of computer bids really did correspond to the Beta distributions from which they had been drawn, and a two-sample t-test confirmed this, finding no significant differences ($p = 1$ for both comparisons) between mean computer-bids for each of the rewards within the same group when pooling across sessions (fig. 4.14b). Similarly, we found no significant differences in computer bids between the 0.2ml and 0.4ml rewards (group 1) or between the 0ml and 0.1ml rewards (group 2) for any of the sessions taken individually. Therefore, any differences in the monkey's bids for different rewards within the same group could not arise due to differences in the computer's bids for those rewards.

(a)



(b)

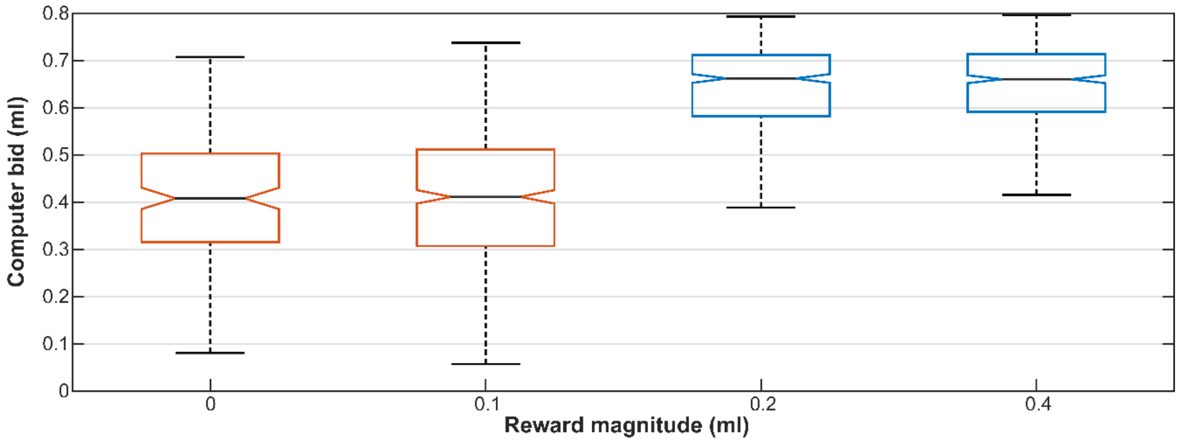


Fig. 4.14) a) The probability density functions (PDFs) of the 'peaked' Beta distributions used to generate computer bids for group 1 (blue) and for group 2 (red). **b)** The realised distribution of computer bids for each of the rewards, pooling across all sessions. Limits of the boxes show the inter-quartile range; medians and confidence intervals for the median are shown by the central line and notches respectively; whiskers show the range of data-points. Note that the 0.4ml and 0.2ml rewards are in group 1 (blue) and the 0.1ml and 0ml rewards are in group 2 (red).

Indeed, the monkey's bids could be used to discriminate between the 4 rewards used in this task, including between those rewards that had their computer bids drawn from the same distribution – and this was true at both the population and session levels.

Pooling data from across all 6 sessions, a 1-way ANOVA found a significant effect of reward magnitude on the monkey's bids [$F(3,1681) = 388.96$, $p = 7.66 \times 10^{-192}$], with a

Bonferroni-corrected multiple-comparisons test showing significantly different mean monkey bids for all 4 of the rewards. This was further supported by a Spearman's rank correlation, showing a significant positive relationship between the reward magnitude and the monkey's bid (Rho = 0.63, $p = 9.5 \times 10^{-189}$), which survived the exclusion of any one of the reward groups.

More importantly, these results could be replicated at the level of individual sessions, with 1-way ANOVAs showing a significant effect of reward magnitude on the monkey's bid in individual sessions (fig. 4.15). A Bonferroni-corrected multiple-comparisons test revealed significant differences between bids for all reward magnitudes in sessions 2, 3, and 4, with significant differences between bids for all reward magnitudes except between those for the 0.4ml and 0.2ml rewards in sessions 1 and 6. Only in session 5 did the monkey's bids fail to show a significant difference for rewards in the same computer-bid group.

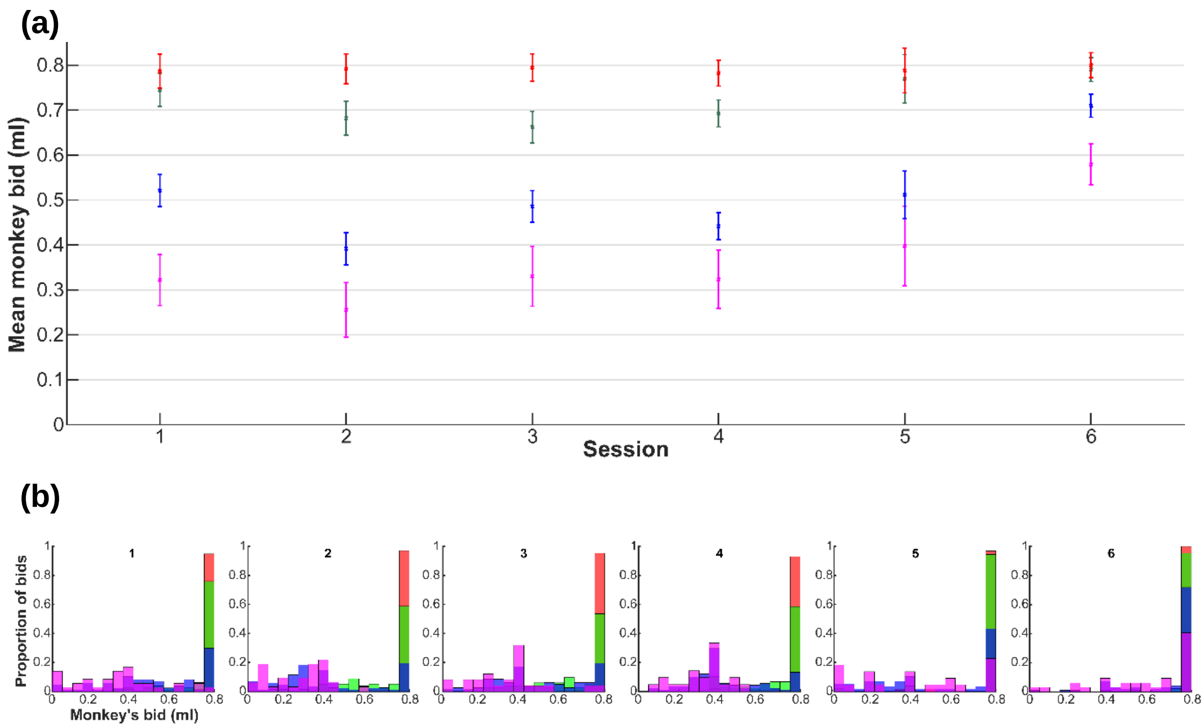


Fig. 4.15) a) Mean bids for the 0.4ml (red), 0.2ml (green), 0.1ml (blue) and 0ml (magenta) reward magnitudes over 6 sessions of the [4R-BDMp] task, using a budget volume of 0.8ml and two different computer-bid distributions. Error-bars are 95% confidence intervals for the mean. **b)** A probability histogram of the data for each session shown in (a).

Of immediate interest is the fact that although the 0ml reward is consistently the lowest valued of the 4 rewards, mean bids for that reward are still non-zero in all sessions, which must be sub-optimal. We could only speculate at this stage as to the reasons for this, though it is worth noting that with a value of zero, the ECM for a bid of 0.4ml, where the maximum bid is 0.8ml, and with a (4,4) beta distribution of computer bids, is only ~0.14ml on each trial. It is possible that this ECM is lower than the cost of determining and making an optimal bid, however, the bids for the 0ml reward in session 6 are significantly higher than the starting position bid of 0.4ml, which suggests that effort was expended in that session in a manner that would only increase the monkey’s losses.

It is therefore possible that the monkey found this condition to be confusing, and past experience in the lab using stimuli that predicted no reward suggests that they can be problematic: including, for example, frustration and non-cooperation of the animal.

Despite these concerns, we could still conclude from these sessions that differences in the monkey’s bids when using different computer-bid distributions for different rewards could not be accounted for by a simple matching of the computer-bid. This was an important result, as it meant that we could now utilise different computer-bid distributions to accelerate the learning of the task by introducing greater costs for the same deviations from optimal behaviour.

In fact, Spearman’s rank correlations on individual session data already showed a significant positive relationship between the reward magnitude and the monkey’s bids in individual sessions (Table 4.7) which was much stronger than had been observed in the preceding BDM sessions. Nevertheless, we were still unable to conclude that the use of more specifically punishing, ‘peaked’, computer bid distributions accounted for this improvement in performance, as the monkey’s experience in the task had also increased over this same period – we returned to this question at a later point (Ch. 5.4).

Table 4.7) Summary statistics for bids, and the Rho and p-value of Spearman's rank correlations between reward magnitude and the monkey's bid, are given for each of the 6 sessions shown in figure 4.15. For each reward magnitude, the mean (M) and standard deviation (SD) of bids are given in millilitres, as is the overall performance in the session, defined as those trials in which a touch-key error was not made.

Session	0.4ml M (SD)	0.2ml M (SD)	0.1ml M (SD)	0ml M (SD)	Rho	p-value	Performance
1	0.79 (0.06)	0.74 (0.11)	0.52 (0.24)	0.32 (0.21)	0.70	3.2×10^{-43}	62.1%
2	0.79 (0.05)	0.68 (0.17)	0.39 (0.25)	0.26 (0.14)	0.76	6×10^{-57}	84.3%
3	0.79 (0.02)	0.66 (0.19)	0.49 (0.22)	0.33 (0.19)	0.69	1.4×10^{-44}	90.6%
4	0.78 (0.07)	0.69 (0.15)	0.44 (0.19)	0.32 (0.13)	0.72	8×10^{-49}	96.2%
5	0.79 (0.07)	0.77 (0.13)	0.51 (0.29)	0.40 (0.30)	0.63	1.2×10^{-23}	75.2%
6	0.80 (1×10^{-30})	0.79 (0.05)	0.71 (0.17)	0.58 (0.24)	0.49	3.5×10^{-19}	97.7%

After the monkey had completed these 6 [4R-BDMp] sessions, we made two key modifications to the BDM task. First, given our concerns with regards to the 0ml reward condition, and the lack of engagement of the animal in these trials (which was often as overt as turning away from the screen in the primate-chair), we decided to continue without a 0ml reward, and focused on seeking further improvements in the monkey's overall performance using rewards which the monkey found engaging and motivating.

Secondly, we were concerned that the monkey was still making too many bids to the very top of the budget-bar for all rewards, and noted the paucity of bids being made to positions that were lower than the bid-marker start position at 0.4ml (fig.4.15b). We speculated that this could be due to a preference for a forward direction of movement of the joystick, and if this was the case we could make better use of the bidding range by starting the bid-marker at the very bottom of the budget-bar.

We tested these two changes in a further set of 6 sessions (fig. 4.16) that otherwise used identical parameters to those described for the preceding 6 sessions, including a (4,4) computer-bid distribution for the 0.1ml reward, and a (8,2) computer-bid distribution for the 0.2ml and 0.4ml rewards – [3R-BDMp].

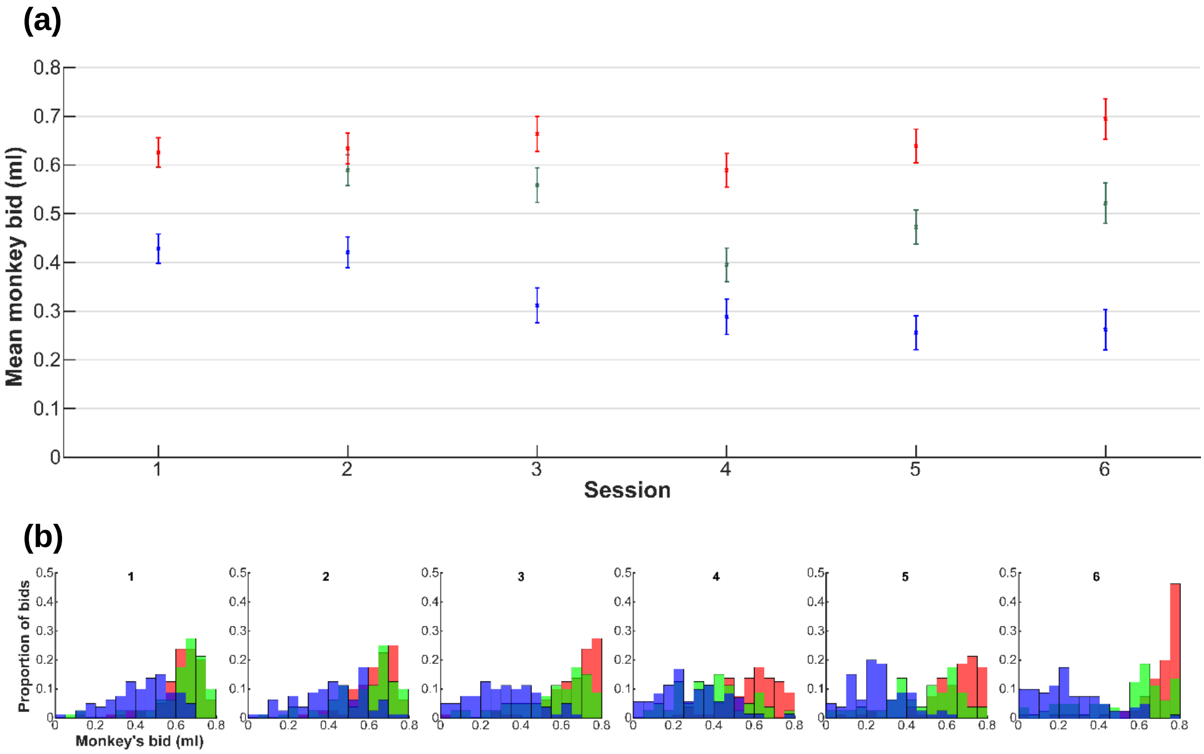


Fig. 4.16) a) Mean bids for the 0.4ml (red), 0.2ml (green), and 0.1ml (blue) reward magnitudes over 6 sessions of the [3R-BDMp] task, using a budget volume of 0.8ml, two different computer-bid distributions and with the bid-marker starting at the bottom of the budget-bar. Error-bars are 95% confidence intervals for the mean. **b)** A probability histogram of the data for each session shown in (a).

The results of Spearman's rank correlations between the reward magnitude the monkey's bid for individual sessions show that the relationship between these two variables had a similar strength to that seen in the preceding session-set after the first two sessions (Table 4.8). Moreover, a 1-way ANOVA on the monkey's bids for showed a significant effect of the reward magnitude in every session, and, for sessions 3-6, a Bonferroni-corrected multiple-comparisons test showed significant differences between bids for all reward magnitudes (in both sessions 1 and 2 the bids for the 0.4ml and 0.2ml rewards were not significantly different p = 1, and p = 0.15 respectively).

Compared to the preceding [4R-BDMp] task, using a starting position for the bid-marker at the bottom of the budget-bar led the monkey to use more of the bidding range (fig. 4.16b). And, after the first two sessions in which the animal may have had to adjust to the new starting position, bids for different rewards were more consistently statistically distinguishable than when using the central bid-marker starting position (shown in fig. 4.10) – particularly for the 0.4ml and 0.2ml rewards, whose bids were consistently concentrated at the top of the bidding range in the [4R-BDMp] condition.

Table 4.8) Summary statistics for bids, and the Rho and p-value of Spearman's rank correlations between reward magnitude and the monkey's bid, are given for each of the 6 sessions shown in figure 4.16. For each reward magnitude, the mean (M) and standard deviation (SD) of bids are given in millilitres, as is the overall performance in the session, defined as those trials in which a touch-key error was not made.

Session	**0.4ml M (SD)**	**0.2ml M (SD)**	**0.1ml M (SD)**	**Rho**	**p-value**	**Performance**
1	0.63 (0.12)	0.63 (0.14)	0.43 (0.15)	0.52	2.4×10^{-18}	87.6%
2	0.63 (0.10)	0.59 (0.15)	0.42 (0.17)	0.52	1.1×10^{-17}	91.6%
3	0.66 (0.14)	0.56 (0.18)	0.31 (0.16)	0.67	6.9×10^{-33}	83%
4	0.59 (0.13)	0.40 (0.17)	0.29 (0.16)	0.63	1.5×10^{-26}	74%
5	0.64 (0.14)	0.47 (0.19)	0.26 (0.14)	0.71	2.4×10^{-38}	80%
6	0.69 (0.15)	0.52 (0.22)	0.26 (0.18)	0.70	2.2×10^{-36}	94.1%

Having established that the BDM could be used to reveal the monkey’s preferences, and that the monkey was not simply matching the computer’s bids, we shifted our focus to training the monkey in a binary choice task so that we would have a measure of value with which to compare the BDM bids. This in turn would help us to continue refining the BDM task as we could, for example, choose reward magnitudes and juices whose values made better use of the bidding range. The testing of alternative task designs and their refinement through changes to various task parameters are described in the next chapter.

5

BDM task testing and refinement

This chapter describes the training of several different versions of the BDM task as well as the accompanying binary-choice-bundle (BCb) task. The results presented here are from experiments conducted in the period between the initial training of a preliminary version of the BDM in the first monkey, Ulysses (Chapter 4), and the final version of the BDM that would be tested in both monkeys (Chapters 2 and 3).

By this point, Ulysses had learnt to place higher bids for larger juice-rewards in a preliminary version of the BDM task. As long as Ulysses' utility for the juice-rewards was described by a monotonic function (i.e. if he perceives more as better), the bids he placed will have reflected his order of preference for the different volumes of juice-reward offered. However, to confirm this, Ulysses would have to be trained in a binary choice (BC) task (Ch. 5.1), in which he would have to choose between a juice-reward and different volumes of the water-budget. Ulysses' choice patterns could then be used to infer a value for that juice-reward in terms of the water-budget, which acts as a numeraire.

But this simple BC task would only come to play an intermediate role in the monkey's training, allowing us to gradually introduce new behavioural requirements for joystick control, but primarily acting as a stepping stone towards a more complicated binary-choice-bundle (BCb) task, which provides a more faithful analogue of the BDM (Ch. 5.3).

While the simple BC task allows us to establish the order of preference for different juice-rewards, it does not reveal their equivalent BDM water-budget values. This is because, in the BDM, the monkey does not choose either some amount of water-budget or the reward, but rather chooses how much water-budget to sacrifice for the juice-reward. Effectively, the monkey uses their bid to choose the minimum amount of water-budget that they will accept alongside the reward, and the monkey will always keep at least whatever water-budget they did not use to bid[*]. Framed in another way, the monkey chooses between the entirety of their water-budget and some amount of water-budget alongside the juice-reward – which is exactly the choice that they are given in the BCb task.

---

[*] The maximum possible cost on a given trial is approximately equal to the monkey's bid (consider the case in which the computer bids infinitesimally less than the monkey). In this case the monkey gains the juice-reward as well as keeping the difference between the maximum of the water-budget and their bid.

During Ulysses' training in the preliminary version of the BDM, a second monkey, Vicer, was being habituated for BDM training also. Due to practical constraints and the delayed training of the second monkey, the monkeys did not always undergo training in the same order as each other, or necessarily take part in the same tasks.

After both monkeys had learnt the BCb task, there was a short period of training in a discrete version of the BDM (Ch. 5.2), but early measures of performance in this task were not promising. We therefore switched back to the continuous paradigm as in the preliminary version of the BDM. Vicer had less experience in the BDM than Ulysses, and needed more training in the continuous version of the BDM at this stage. Therefore, we used this time to explore the effects of several different task manipulations in Ulysses alone: first, we tested the effects of manipulating the computer-bid distribution (Ch. 5.4), and later we went on to change the starting position of the bid-marker (Ch. 5.5).

Ultimately, however, several simple changes to the task parameters, initially tested in Vicer, led to the changes in task design upon which we would base the final version of the BDM task (Ch. 5.6).

Different task designs required several preparatory sessions in which the monkeys were introduced to new task stimuli or trained with new behavioural requirements – with the relevant training tasks being described in Chapter 4. For example, the monkeys often had to be trained in target-tasks between different versions of the BDM, and this was necessary to improve their bidding precision after long periods of absence from BDM training (due to implantation surgery[*], for example). Similarly, BCb tasks would often have to be used to reassess the monkeys' preferences and values after changes of volume or flavour of the water-budget and juice-rewards. To avoid repetition, we have not described every instance of such testing in these supplementary tasks, but rather describe when and why they were used and refer to their original descriptions.

Some of the data-sets included in this chapter are incomplete, and in some cases, we only collected a few sessions before deciding to abandon an experiment for

---

[*] However, it should be noted that the monkeys were not head-restrained for any of the experiments presented in this thesis. The implantation surgery was made for prospective neurophysiological recording experiments, and was conducted well in advance of these experiments to allow for better bone growth around the implant.

some other version of the task. Ideally of course, we would have spent more time on these various possibilities, to more clearly characterise the effects of a given manipulation on the monkeys' bidding behaviour. However, due to a combination of time constraints and the exploratory nature of many (if not all) of our manipulations, we had to be flexible in switching between different possible methods, and couldn't afford to spend too long on manipulations that didn't present a clear benefit to the monkeys' performance within a few weeks of testing.

The sections of this chapter are presented according to the order in which the different task-types were trained in the first monkey. The order in which training progressed was the same for the second monkey, Vicer, except for the exclusion of the tasks described in sections 5.4 and 5.5. After testing of the last version of the task that is shown in section 5.6, the monkeys were both trained in the final version of the task - discussed in Chapters 2 and 3.

5.1 - Binary choice testing and new joystick controls

To refine the BDM task, we needed to select rewards and budget volumes that would allow the monkey to make use of as much of the bidding range as possible. Moreover, we wanted to be able to assess the monkey's valuations of reward juices in terms of the water-budget in the BDM task to see whether bids were meaningful beyond a simple ordinal ranking. Ultimately, we would seek to compare BDM bids to the values inferred from a BC task which compared a bundle of some water-budget and juice-reward with the full water-budget volume alone (Ch. 2.3).

However, before introducing the monkey to a complex task involving bundles of rewards, we first trained them on a simpler BC task offering choices between various volumes of water-budget and different reward-predicting fractals. This simple BC task would allow us to train the monkey towards the harder bundle version of the task, as well as giving us some measure of the relative values of different rewards to the budget. We could also use the monkeys' behaviour in the BC task to confirm that they understood the reward and budget predicting stimuli, as our only previous indication of this outside the BDM task was from an analysis of the lick-data for reward-predicting fractals.

One of the key issues that we faced in the earliest versions of the BDM task was a propensity for bids to be concentrated at the extremes of the bidding range (Ch. 4.2), and it was possible that this was in part due to some of the rewards being valued more than the maximum water-budget volume. The BC task would therefore help us to select rewards that were within the water-budget range. However, we also suspected that control of the joystick and motion of the bid-marker were contributing to this phenomenon of maximal bidding.

With few controls on the monkey's movement of the joystick, as well as such a short choice epoch in the preliminary BDM task (fig. 4.10), it was possible that the monkey was following some simple heuristic strategy, or placing bids without paying sufficient attention to the bid-marker position. As animals tend to trade-off accuracy for speed under time constraints[78], we reasoned that an increase in the choice time (from 2s to 4s) would reduce the pressure to act quickly and could support the learning of more precise bidding behaviour.

Furthermore, the monkey had previously been free to move the joystick in the ‘Offer’ epoch before bids were to be placed, and could therefore prepare a movement and move the bid-marker without continuing to attend to on-screen events. This issue was compounded by the fact that the monkey didn’t have to select a bid in the preliminary version of the BDM task, rather, the final position of the bid-marker was taken as the monkey’s bid. Together, these factors meant that the monkey could place a bid without attending to the task (for example, on some occasions, we noticed that the monkey would push the joystick forwards in the ‘Offer’ epoch and then look away from the screen, and could then still collect any reward/budget at the end of the trial).

Therefore, as well as increasing the choice time, we introduced two new behavioural requirements relating to joystick control. First, the monkey would now have to hold the joystick in a central position (within a tolerance-window of 2% of the maximum voltage deflection magnitude in each direction – see Appendix 2) for the duration of the ‘Offer’ epoch, ensuring that they would have to attend to the timing of the ‘Choice’ epoch, and thus initiate their movement in a more controlled manner, otherwise the trial would end with an error (not-centred error). Secondly, a bid (or choice in the BC task) was now only recognised if the bid/choice-marker had been stabilised for 250ms before the end of the choice epoch, and once it had remained static in some position for this amount of time it could no longer be moved. This meant that if the joystick/marker was still moving by the end of the choice-epoch, the trial would end with an error (no-choice error).

The monkey was trained in a simple version of the BC task (fig. 5.1) using these new behavioural requirements and an increased choice-time.

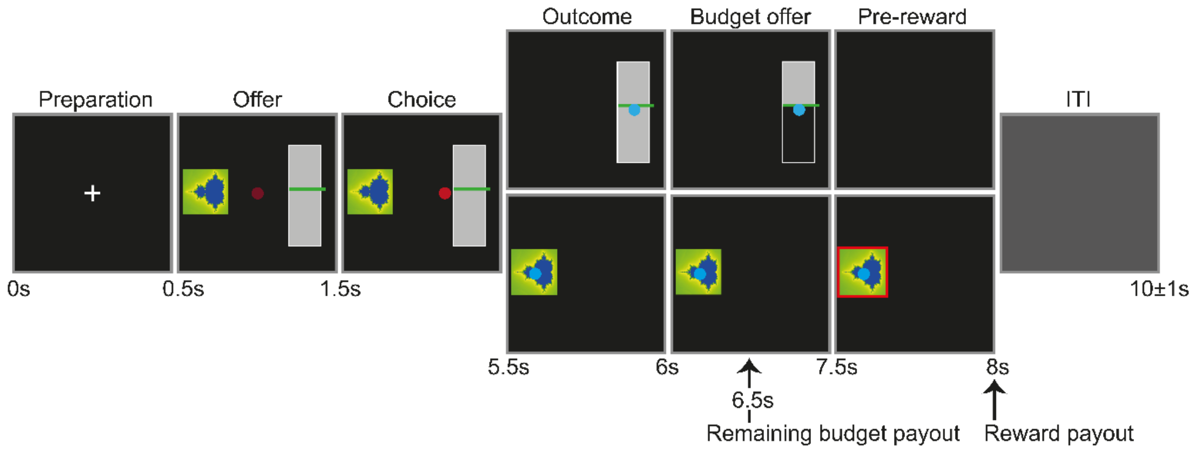


Fig. 5.1) Trial structure of the simple binary choice (BC) task. Following a brief preparation period to orient the monkey to the screen, the two reward objects that monkey could choose between were shown for 1s (the side on which the budget/reward appeared was randomised). During this 'Offer' epoch the monkey had to maintain hold of the joystick and keep it centrally positioned, else the trial would end with an error. At the beginning of the 'Choice' epoch the choice-marker changed colour to a brighter shade of red, indicating that it could now be moved left/right using the joystick. The monkey had to bring the marker to rest by the end of the 4s choice epoch by returning the joystick to its centre position. Once the choice-marker was stabilised within the chosen stimulus, it changed colour to cyan, indicating that a choice had correctly been made. At the end of the 'Choice' epoch, the outcome of the choice was displayed for 0.5s, with the unchosen stimulus disappearing at this point. Budget/reward liquids were then paid out either in the 'Budget' or at the end of the 'Pre-Reward' epoch, depending on the monkey's choice – stimulus changes in these epochs, and the time of delivery of reward relative to those stimulus changes were kept consistent with the BDM task. The final 'Pre-reward' epoch was followed by a variable inter-trial interval (ITI) of 2±1s, with random durations distributed by a truncated exponential function. Any error led to a blue error time-out screen of duration equal to remaining trial time + 3s.

We initially trained the monkey for two weeks on the BC task alone, gradually introducing the new behavioural requirements for joystick control. Early in the training of this BC task the monkey showed a significant right-hand-side (RHS) bias, which was remedied by altering the gain of choice-marker movement speed to be unequal for the two directions (Appendix 2). We titrated the necessary ratio in gains by attempting to achieve a close to equal proportion of left and right choices when the two options were the same (we used two of the same fractal stimuli as the reward options in these sessions). A binomial test was used to confirm that RHS choice was not significantly different to a choice probability of 50%.

Results of the BC task in the first monkey:

Once Ulysses had learnt the new behavioural requirements and was moving the joystick without a consistent side-bias, we began testing different volumes of juice-reward against the water-budget. Our original reward set had made use of 0.1ml, 0.2ml and 0.4ml of blackcurrant juice rewards, but early testing in the BC task suggested that the 0.4ml reward was more highly valued than the 0.8ml of water water-budget*. Later testing suggested that 0.1ml, 0.2ml and 0.3ml blackcurrant rewards would make better use of the bidding-range (fig. 5.2).

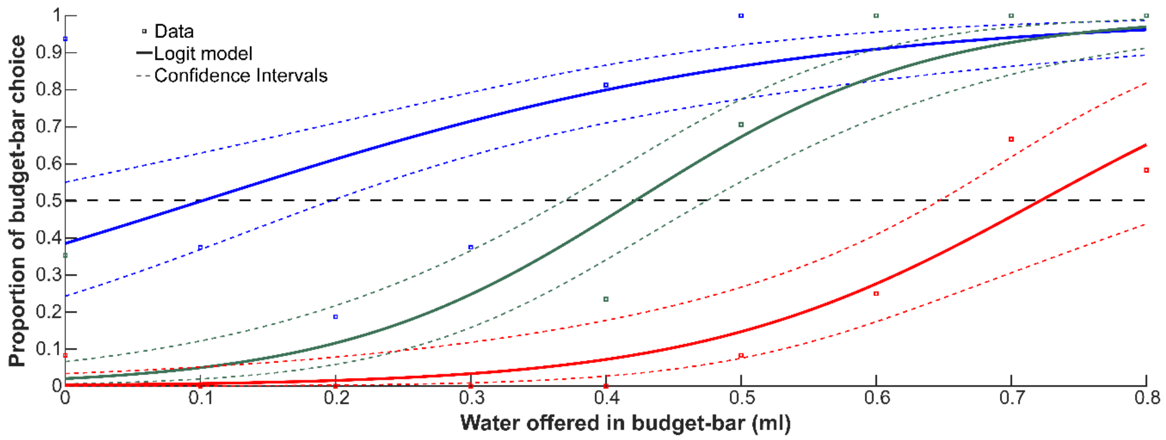


Fig. 5.2) Data from an example session of the simple BC task, showing the choice data and logit-model fit for each of three blackcurrant reward volumes: 0.1ml (blue), 0.2ml (green) and 0.3ml (red). The logit model fit was used to infer the value of the reward in terms of water-budget by finding the point of indifference for each reward (0.5 proportion of budget-bar choice, black dashed line). The 0.1ml reward was valued at 0.1ml, the 0.2ml reward at 0.42ml, and the 0.3ml reward at 0.72ml of water-budget.

Values of the blackcurrant reward in terms of water-budget were inferred from the choice data by fitting a logistic regression model to the proportions with which the budget-bar stimulus was chosen at every different volume of water that had been offered in it (Appendix 3). Different water volumes were compared with a given juice-reward in steps of 0.1ml of water. The water volume at which the subject was

* This could explain the large number of maximal bids that were placed for this reward in the preliminary BDM task (fig. 4.13 and fig. 4.15).

indifferent between the two rewards, i.e. when the proportion of budget-bar choice was 0.5, indicated the value of that volume of blackcurrant juice in terms of water.

Again, the values inferred from this BC task were not expected to be equivalent to the BDM bids, but nevertheless provided our best approximation of the rewards' values, and would therefore provide a rough guide for calibrating juice volumes to be used in the BDM task until the monkey had been trained in the more complex bundle-version of the BC task (Ch. 5.3).

Having changed the reward set to the 0.1, 0.2 and 0.3ml blackcurrant rewards, we now reintroduced the BDM task also using the new, longer, 'Choice' epoch time of 4s and the centre and stabilisation requirements for the joystick, [0.8-BDM-BC]. The monkey was trained concurrently in the BC and BDM tasks, with alternating blocks of 8 trials each (1 trial for each of the water-budget volumes tested in the BC task, ranging from 0.1ml to 0.8ml in 0.1ml steps), for 13 sessions (fig. 5.3).

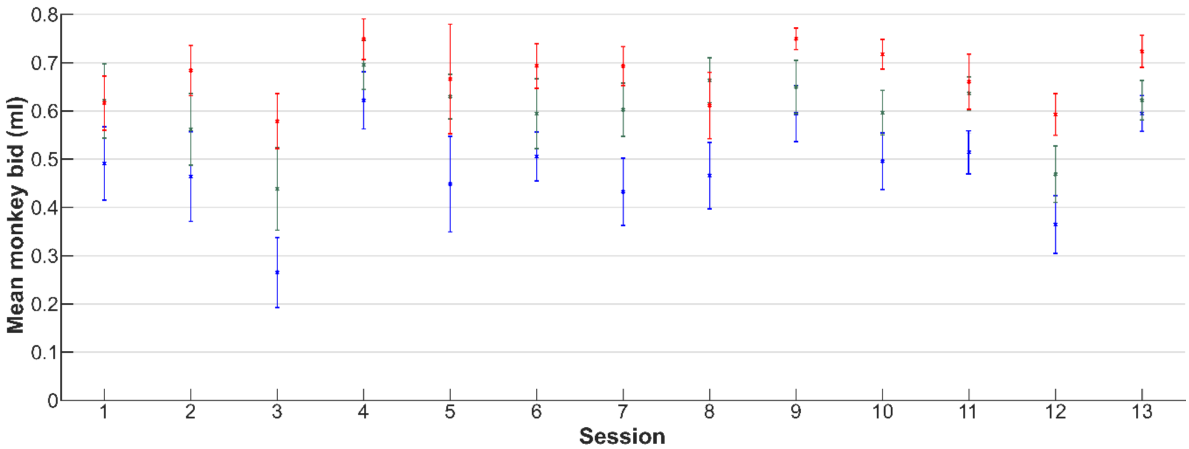


Fig. 5.3) Mean bids for the 0.3ml (red), 0.2ml (green), and 0.1ml (blue) reward magnitudes over 13 sessions of the [0.8-BDM-BC] task (with interleaved BC trials, new behavioural requirements for joystick control, and a choice time of 4s). Error-bars are 95% confidence intervals for the mean.

By alternating between the BC and BDM tasks we could maintain the monkey's performance in both tasks. We had also hoped that we could track changes in the values of rewards across sessions due to satiety, for example. However, behaviour in both the BC and BDM tasks was still too variable, with a high degree of uncertainty in the inferred values precluding such an analysis: for example, the mean

95% confidence interval range for the inferred value of the 0.2ml reward from the 13 sessions tested was ~0.18ml, or ~23% of the total bidding range.

Moreover, alternating between the tasks meant that we had approximately half as many BDM trials per session, thus, inferred values in the BDM task were also much more variable, and performance - as measured by the correlation between monkey's bid and the reward volume (Table 5.1) - was significantly worse than in the preceding version of the BDM: The Spearman's Rho values of this correlation were significantly greater in the preceding version of the task (M = 0.62, SD = 0.086) than for the current version (M = 0.46, SD = 0.11, p = 0.003), as determined by a two-sample t-test*.

The proportion of 'no-choice' errors in this combined BDM and BC task also reflected the values of the rewards on offer. On average, across the 13 sessions, the monkey made a 'no-choice' error (i.e. failed to make a bid in the BDM, or to select an option in the BC task) on 11.5% of trials for the 0.1ml reward, 4.9% of trials for the 0.2ml reward and on only 2.7% of trials for the 0.3ml reward (fig. 5.4) – with significantly more 'no-choice' errors for the lowest value reward than for the highest value reward ($p = 6.04 \times 10^{-4}$, Bonferroni-corrected multiple-comparisons t-tests).

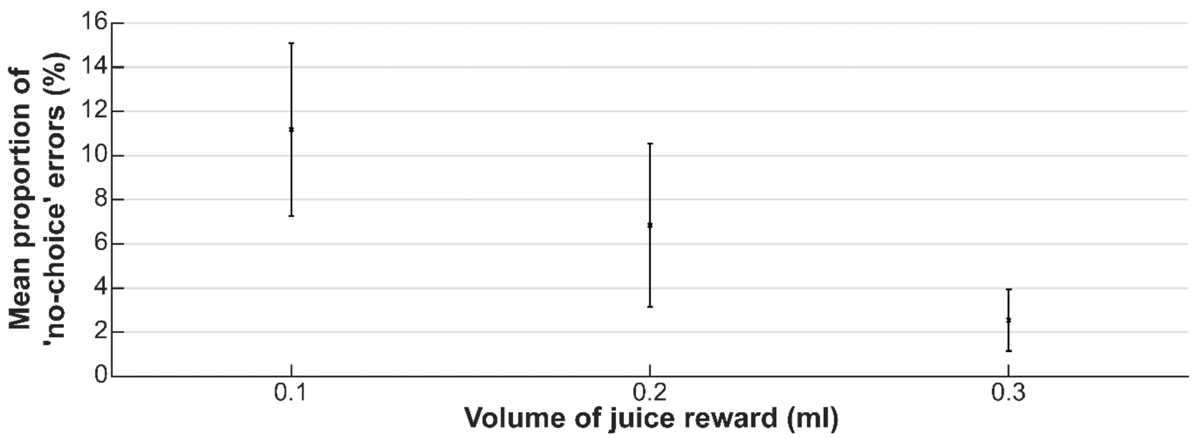


Fig. 5.4) The mean percentage of trials in which a 'no-choice' error was made for each reward volume from across the 13 sessions in the data-set shown in figure 5.3. The monkey made significantly fewer no-choice errors for trials with a greater motivational value.

* Compare Tables 4.7 and 5.1 for the results of the Spearman's rank correlations of reward volume and monkey's bid for individual sessions.

Table 5.1) Showing the mean (M) and standard deviation (SD) of values for each reward as inferred from the BDM and the BC tasks, as well as the Rho and p-values of Spearman's rank correlations of the monkey's BDM bids and the reward volume for each session. The overall performance for each session, across both BDM and BC trials is also shown.

Session	BDM 0.3ml M (SD)	BDM 0.2ml M (SD)	BDM 0.1ml M (SD)	BC 0.3ml M (SD)	BC 0.2ml M (SD)	BC 0.1ml M (SD)	Rho	p-Value	Performance
1	0.62 (0.21)	0.62 (0.24)	0.49 (0.26)	0.79 (0.07)	0.72 (0.10)	0.28 (0.12)	0.21	0.011	79.7%
2	0.68 (0.18)	0.56 (0.26)	0.46 (0.26)	>0.8 (N/A)	0.63 (0.08)	0.40 (0.19)	0.39	4.8×10^{-6}	51.7%
3	0.58 (0.19)	0.44 (0.27)	0.27 (0.23)	0.80 (0.003)	0.71 (0.16)	0.30 (0.15)	0.51	1.3×10^{-9}	71.4%
4	0.75 (0.14)	0.70 (0.16)	0.62 (0.20)	0.80 (0.004)	0.59 (0.12)	0.38 (0.09)	0.44	9.6×10^{-8}	85.5%
5	0.67 (0.27)	0.63 (0.14)	0.45 (0.25)	0.80 (0.004)	0.63 (0.08)	0.41 (0.18)	0.47	2.2×10^{-6}	48.9%
6	0.69 (0.11)	0.59 (0.25)	0.51 (0.19)	0.80 (0.004)	0.73 (0.12)	0.42 (0.09)	0.43	4.6×10^{-7}	54.9%
7	0.69 (0.13)	0.60 (0.19)	0.43 (0.22)	0.80 (0.004)	>0.8 (N/A)	0.27 (0.19)	0.50	2×10^{-9}	61%
8	0.61 (0.24)	0.66 (0.16)	0.47 (0.23)	>0.8 (N/A)	0.71 (0.11)	0.41 (0.10)	0.32	1.3×10^{-4}	91.3%
9	0.75 (0.07)	0.65 (0.17)	0.59 (0.18)	>0.8 (N/A)	0.65 (0.10)	0.35 (0.09)	0.51	3.6×10^{-9}	88.9%
10	0.72 (0.09)	0.60 (0.13)	0.50 (0.16)	>0.8 (N/A)	0.70 (0.003)	0.40 (0.003)	0.62	1.5×10^{-11}	93.2%
11	0.66 (0.16)	0.64 (0.10)	0.51 (0.11)	>0.8 (N/A)	0.60 (0.003)	0.42 (0.07)	0.50	1.7×10^{-7}	94.1%
12	0.59 (0.12)	0.47 (0.16)	0.36 (0.14)	>0.8 (N/A)	0.60 (0.07)	0.35 (0.07)	0.60	5.9×10^{-10}	74.6%
13	0.72 (0.09)	0.62 (0.11)	0.59 (0.12)	0.70 (0.003)	0.60 (0.10)	0.33 (0.06)	0.44	3.5×10^{-6}	95.4%

Training of the BC task in the second monkey:

The second monkey, Vicer, was also trained in this simple BC task before progressing to the more complex bundle version of the task. As we had already established that the BDM could be used to infer the preferences of the first monkey (Ch. 4.2), we could now afford to invest time in the BC task before training the second monkey in the BDM - allowing us to refine our selection of reward and budget juices and volumes, as well as several other task parameters, so that the BDM training in the second monkey could progress in a way that would help us to avoid some of the problems and confounds that we faced in the first monkey (most notably, not being able to confirm whether the budget-range was appropriate for the rewards being bid for).

Unlike the first monkey, the second monkey did not undergo concurrent training of the BDM and BC tasks, as they had not yet been trained in the BDM task. Instead, for Vicer, the BC task served only as an intermediate task before training of the BCb task and as a confirmation that learning of stimulus values had taken place.

After conducting Pavlovian stimulus learning (PSL) sessions for the fractal and budget-bar stimuli, as described for the first monkey (Ch. 4.1), we confirmed that the second monkey had understood these stimuli by training them in a BC task with the same structure as described above (fig. 5.1), but with both choice options being different volumes of budget or different volumes of reward. Thus, these sessions acted as simple tests of stimulus learning, as we already knew that the monkey should always choose the higher volume option whenever the two options were of the same juice type (that is, one of the options was always strictly better).

We also used this early BC training as an opportunity to introduce Vicer to the touch-key holding, joystick centring, and joystick stabilisation behavioural requirements. Furthermore, Vicer’s training in the BC task would prepare him for the more complex joystick control required for accurate performance of the target-task (Ch. 4.1), and ultimately the BDM, by introducing him to the visuomotor contingencies of moving an on-screen marker in a simpler setting.

Testing different fractals, and therefore juice rewards, against one another in 4 sessions of the BC task suggested that the monkey had learnt the values indicated by the fractals: he reliably chose the higher value option in 96.4% of 550 correctly completed trials (choosing the higher value option on 90.4% of trials in the worst of these sessions).

Moreover, these sessions also revealed a slight right-hand-side (RHS) bias in the monkey's choices, as he chose the RHS option significantly more often when choosing the worst of two options (15 RHS choices vs. 5 left-hand side, LHS, choices) and when the two options were the same (150 RHS choices vs. 96 LHS choices), as assessed by a binomial fit to the choice data, using an alpha level of 0.05 (Appendix 3). While the strength of this bias was not sufficient at this stage to greatly influence the monkey's pattern of choices, a RHS bias could become a problem at later stages of training.

After testing the monkey's choices using two fractal juice-reward options, we moved on to testing the monkey's choices using two budget-bar stimuli in 8 sessions of the BC task. The monkey quickly learnt to reliably choose the higher value budget-bar, choosing the higher volume option on more than 90% of trials in all sessions after the first (fig. 5.5). Interestingly, there was also no indication of any side-bias across these sessions – overall there was no significant difference between LHS and RHS choice when the two options offered the same volume of water-budget (103 LHS/94 RHS choices), or when choosing the lower value option (71 LHS/57 RHS choices).

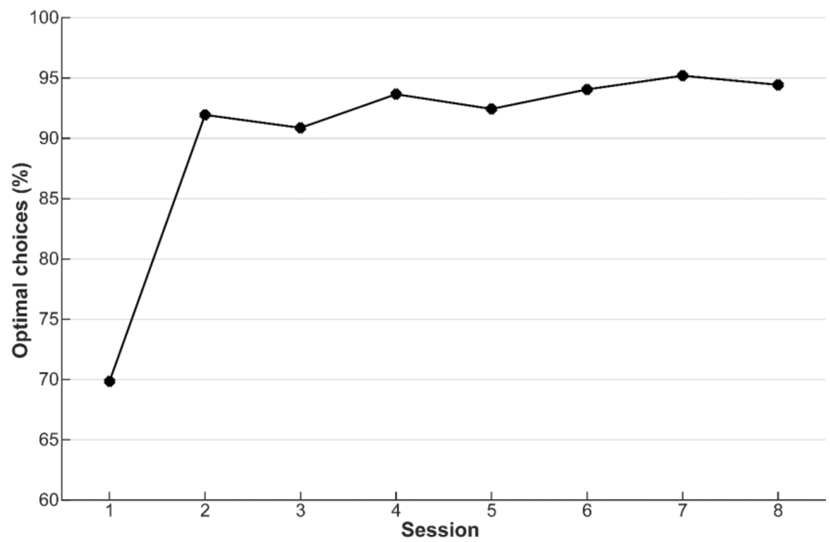


Fig. 5.5) Percentage of trials on which Vicer chose the higher volume option over 8 sessions of the BC task using two budget-bar stimuli.

5.2 - BDM task with discrete bids

Having seen little to no improvement in the correlation between bids and reward magnitude over several iterations of the BDM task, we considered the possibility that a simpler, discrete, bidding task, [0.8-DBDM-BC] could yield better results by presenting the monkey with a more tractable problem. We reasoned that a version of the task with a limited number of discrete bids could lead to an improvement in the monkey's performance by reducing the number of possible options - this could help the monkey to track the payoffs for each given possible bid more easily by learning the expected payoffs for each of the specific possible bids, rather than having to abstract a relationship between bid position and payoff on a continuous budget-bar. Thus, the task may be considered to more closely resemble a multi-armed bandit task, which monkeys have been able to learn in several different experiments.

The [0.8-DBDM-BC] task structure (fig. 5.6) did not differ significantly from that used in previously trained versions of the BDM (fig. 4.10), and made use of the same timings, budget, and rewards as had been used in the preceding [0.8-BDM-BC] task (which included an interleaved BC task). The only significant differences were the segmentation of the budget-bar into 8 equal parts, each representing 0.1ml of water water-budget, and, an increase in the size of the monkey and computer bid-markers to cover one of these segments entirely. Moreover, the monkey could place a bid of 0ml by moving the bid-marker below the lowest segment.

Although both monkeys could place bids in this discrete version of the task, there was little differentiation between bids for the different reward volumes.

For the first monkey, although a 1-way ANOVA found a significant effect of reward volume on bids for all but 2 sessions (sessions 2 and 9, fig. 5.7a), a Bonferroni-corrected multiple-comparisons test only found a significant difference in mean bids for all three groups in a single session, session 5. Otherwise, a difference in mean bids between groups could only reliably be found at a population-level, when pooling bids across all 9 sessions (fig. 5.7b). For each session, we also computed the value of Spearman's Rho for the correlation between the monkey's bid and the reward volume, finding a mean Spearman's Rho of 0.43 (SD = 0.14).

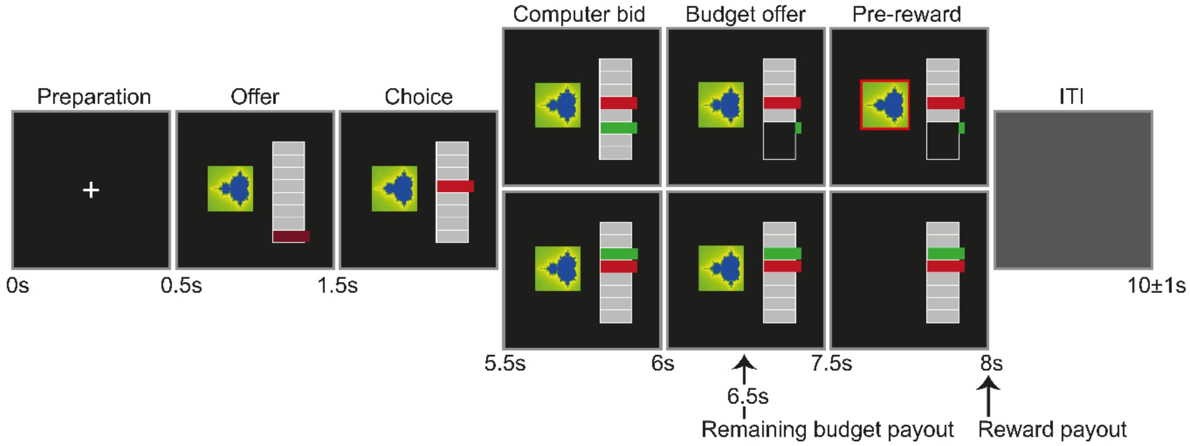


Fig 5.6) Task design for the discrete [0.8-DBDM-BC] version of the BDM. The budget-bar was divided into 8 equal segments, each representing 0.1ml of water, and the bid-marker now covers a whole segment of the budget-bar. When the monkey's bid was equal to the computer bid, the auction was resolved by a coin-flip.

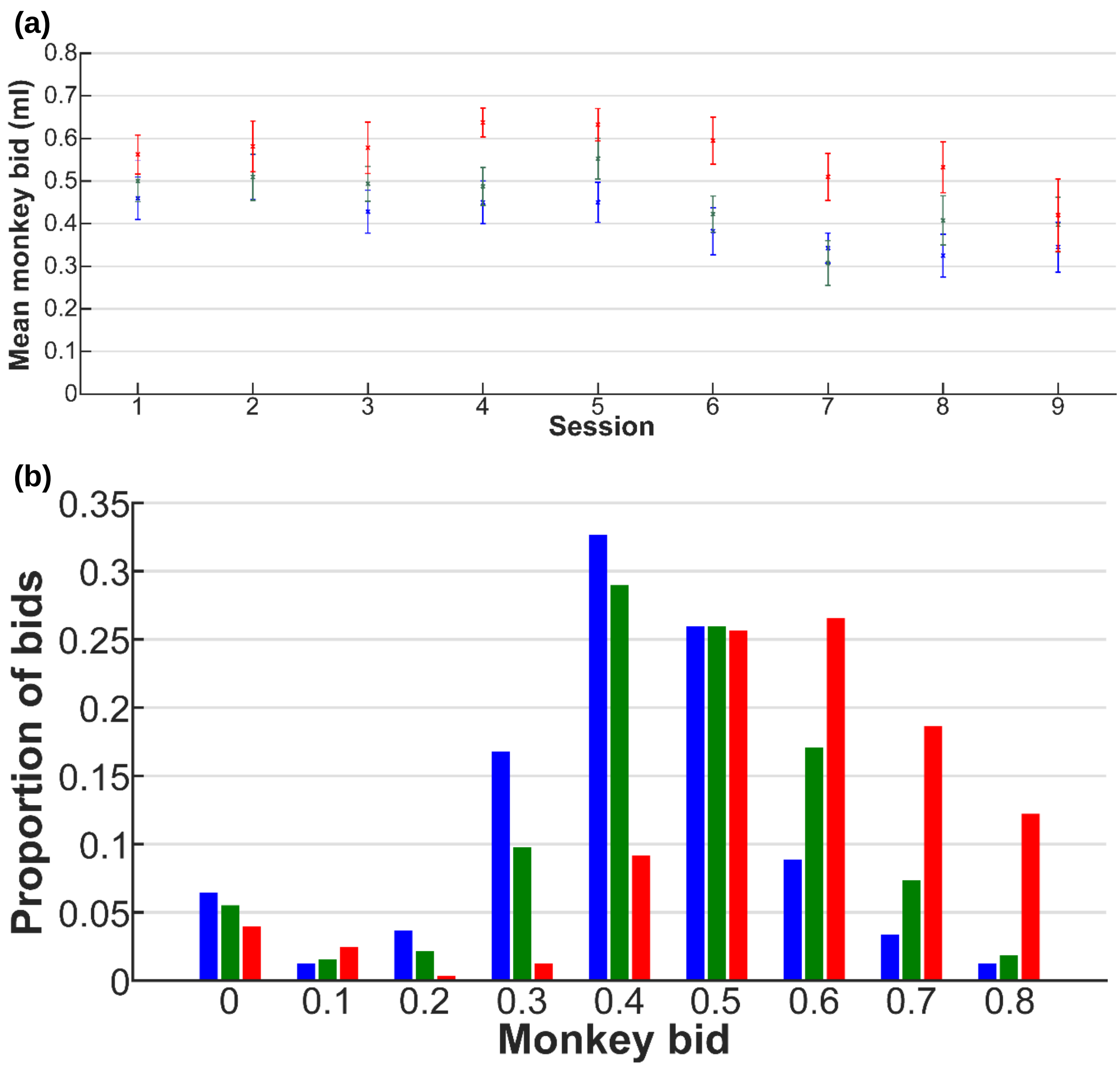


Fig 5.7a) The mean bids for the first monkey, Ulysses, across 9 sessions of the [0.8-DBDM-BC] task. Error bars are 95% confidence intervals of the mean. **b)** The proportion of each possible bid made for each of the three rewards, pooling data from across all 9 sessions. Data in both panels are shown for the 0.3ml (red), 0.2ml (green) and 0.1ml (blue) rewards.

Discrete BDM with Vicer:

The second monkey, Vicer, was also tested using the same discrete BDM task with only a few minor changes. The total budget volume was 0.9ml, not 0.8ml, in his case – having been adjusted according to the values estimated from his values in a preceding BC task. And, according to his performance in a preceding target task, the divisions used represented sixths of the total budget volume, and were therefore 0.15ml each. Unfortunately, as in the first monkey, we only found a weak relationship between the monkey's bids and the reward volume. In a set of sessions using only 2 rewards, [2R-0.9-DBDM-BC], (fig. 5.8) we found a mean Spearman's Rho of 0.29 (SD = 0.06) for the correlation between the monkey's bid and the reward volume, and two-sample t-tests only found a significant difference in bids for the two rewards in sessions 2 and 3.

Furthermore, addition of a third 0.2ml reward option, [3R-0.9-DBDM-BC] led to a collapse in the monkey's performance (fig. 5.9), with a mean Spearman's Rho of only 0.13 (SD = 0.06) for that data set. Even at the population-level, the strength of the correlation between the monkey's bid and the reward volume was only 0.11, suggesting a high degree of inconsistency in the monkey's bids.

The fact that measures of bidding consistency were so poor for both monkeys, and specifically that they represented a decrease in the performance of the first monkey relative to some of the preliminary task versions (Ch. 4.2), led us to abandon the discrete version of the task at this stage.

However, before concluding this section, it is worth briefly presenting the results of some preliminary tests using a touchscreen to make discrete bids.

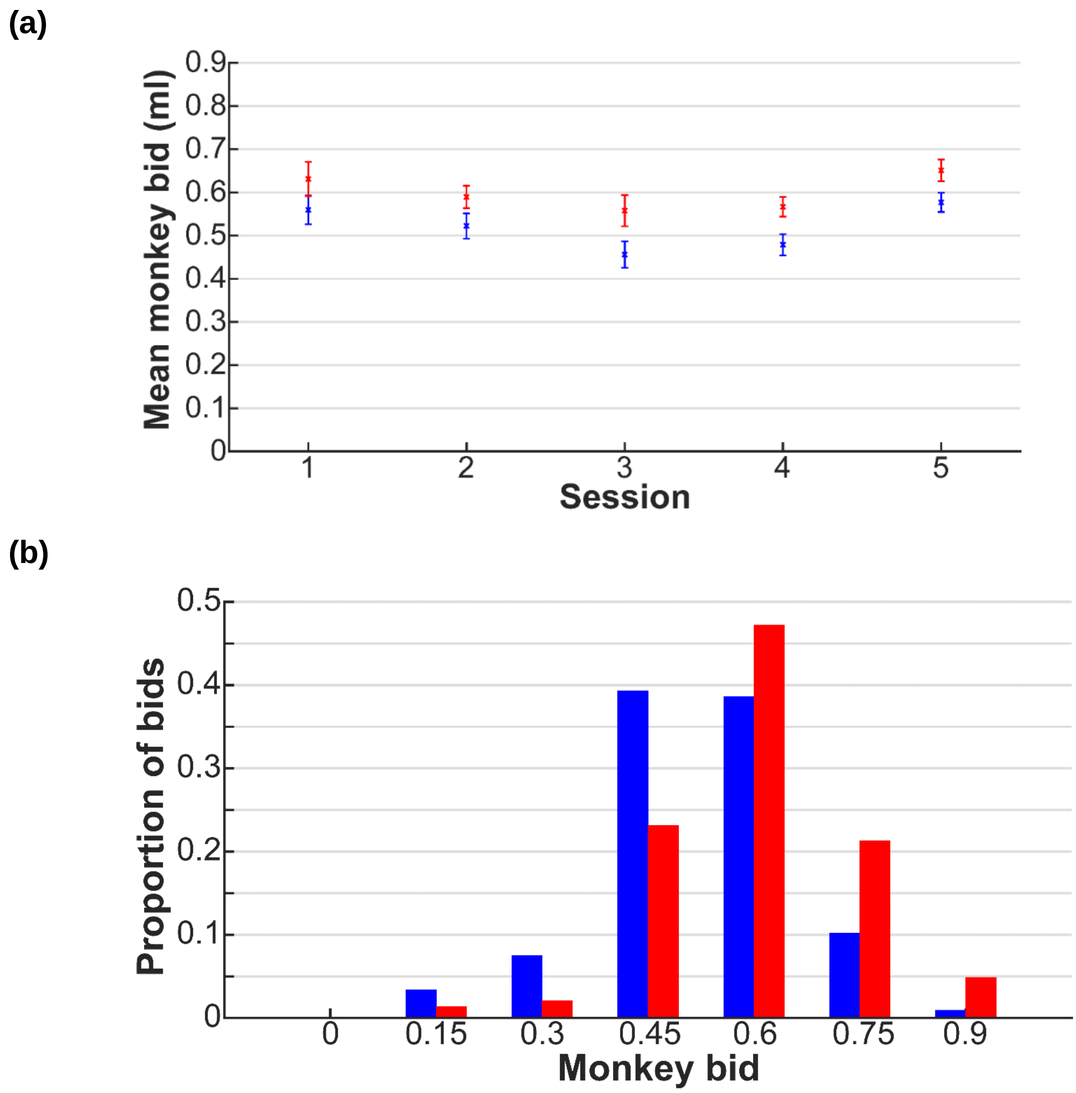


Fig. 5.8a) Mean bids for the second monkey, Vicer, across 5 sessions of the [2R-0.9-DBDM-BC] task, using only 0.1ml (blue) and 0.3ml (red) rewards. Error bars are 95% confidence intervals of the mean. **b)** The proportion of each possible bid made for each of the two rewards, pooling data from across all 5 sessions.

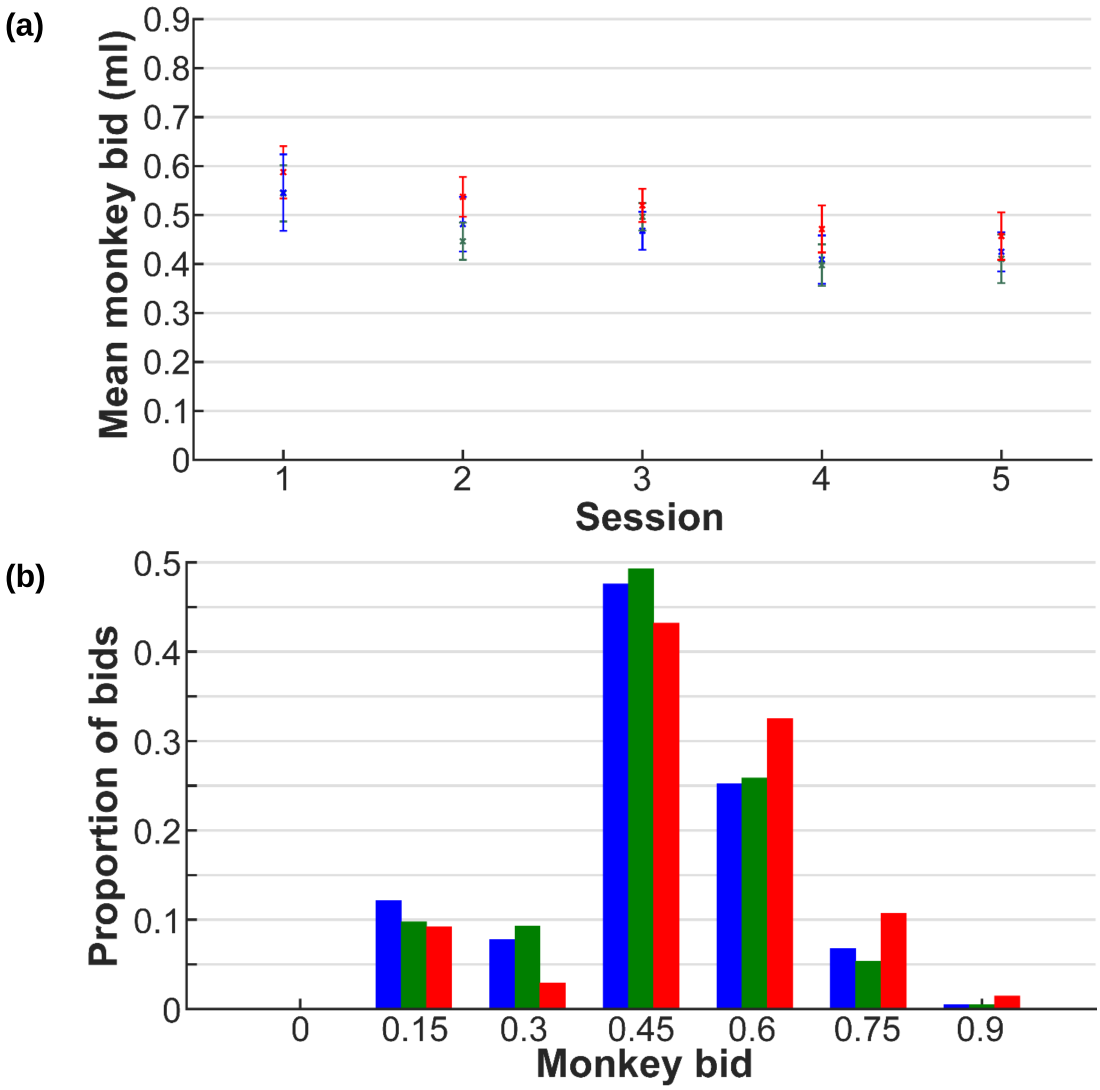


Fig. 5.9a) Mean bids for the second monkey, Vicer, across 5 sessions of the [3R-0.9-DBDM-BC] task, using 0.1ml (blue), 0.2ml (green), and 0.3ml (red) rewards. Error bars are 95% confidence intervals of the mean. **b)** The proportion of each possible bid made for each of the two rewards, pooling data from across all 5 sessions.

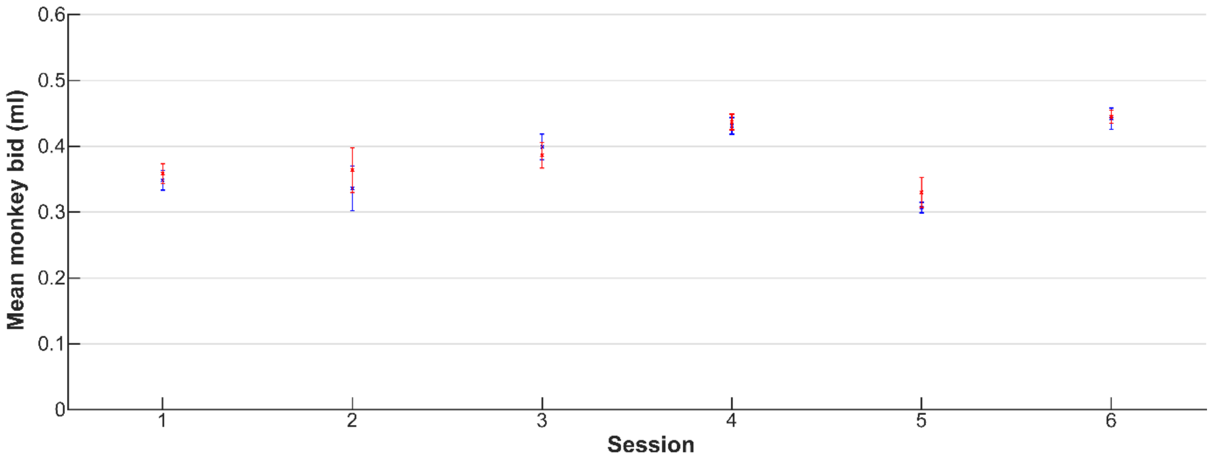
<u>**Discrete bidding with a touchscreen:**</u>

Difficulties in precisely controlling the joystick may have resulted in much of the noise observed in our previously collected bidding data, and we therefore considered the touch-screen as an alternative method for the monkey to make bids, speculating that a direct reaching movement may be simpler and more natural for the monkey than the visuomotor transformation required to make move the bid-marker with the joystick.

The results of a short set of sessions in which the first monkey used a touch-screen, [0.8-TBDM-BC] to make bids are presented here (fig. 5.10). The task and stimuli were displayed as in fig. 5.6, with the exception that no bid-marker was presented in the 'Offer' epoch. Instead, the cross from the presentation epoch stayed on-screen until the end of the offer epoch, signalling that the monkey could now place a bid by touching the budget-bar. The monkey had to then release a touch-key and move to touch one of the 8 segments of the budget-bar, at which point the bid-marker would appear at that location. The monkey was not allowed to touch the screen more than once, and any touch outside of the budget-bar area led to an error time-out followed by the ITI epoch and repetition of the trial.

After the monkey had achieved over 90% overall performance in an associated touch-screen target-task, we moved on to testing the monkey in this touch-screen version of the BDM. Unfortunately, the monkey's bids could not be used to differentiate the two reward volumes tested in any of the 6 sessions, as measured by two-sample t-tests of the bids for each reward.

Due to time constraints, and the lack of any indication that a discrete version of the task would lead to superior bidding (either with the joystick or touch-screen) we had to quickly return to testing the monkeys using the continuous version of the BDM.



5.10) Mean bids of the first monkey for the 0.3ml (red) and 0.1ml (blue) rewards in the [0.8-TBDM-BC] touch-screen version of the discrete BDM task. Error-bars are 95% confidence intervals of the mean.

5.3 - Binary-choice-bundle (BCb) task

Both monkeys had already been trained in a simple binary choice (BC) task, allowing us to confirm that they understood the values associated with each stimulus. The simple BC task gave us estimates of the values of the juice-reward stimuli relative to the water-budget stimulus, and we used these estimates to adjust the volumes, juice types and juice concentrations to make better use of the bidding range. However, the main purpose of the BC task was as an intermediate stage to training the monkey in a more complex bundle-version of the BC task (BCb), which was necessary as the values inferred from the simple BC task cannot be treated as equivalent to the values inferred from the mean bids of the same rewards in the BDM.

For a binary choice task to have a degree of equivalence with the BDM it would need to interrogate the monkey's choices at the same wealth-level. In the BDM, the monkey can be considered to be making a choice between the whole of their budget and a combination of some budget and the reward; their willingness to pay (WTP), indicated by their bid, is the maximum amount of budget that they would sacrifice to gain the reward. Framed in another way, the monkey's bid is the maximum acceptable cost for the reward in terms of water-budget, given that the monkey will retain the remaining water-budget alongside the reward.

Therefore, if the monkey was to pay an amount equal to their bid (e.g. if the computer bid is infinitesimally lower than the monkey's bid), then the amount of remaining water-budget alongside the reward should be indifferent to the entire budget, such that:

$$\boldsymbol{Remaining\ Budget = Budget - Cost}$$

$$\boldsymbol{Utility(Remaining\ Budget + Reward) = Utility(Budget)}$$

This equality between the two utilities is satisfied when the cost is equal to the monkey's bid. Any cost higher than this should lead to preference of the whole budget alone, whilst any cost lower than this should lead to preference of the bundle of some reduced amount of budget and the reward (as the value of the reward would outweigh the cost). The optimal monkey bid is the point at which the utility of the cost and the utility of the reward are equal and opposite when considered alongside the utility of the remaining budget. If the monkey has a typical concave utility function for water, then he should more readily sacrifice a given amount of water for some amount of juice when the choice is between two levels of water higher on the utility function, as opposed to lower on the utility function.

We therefore designed a binary-choice-bundle (BCb) task, in which the monkey chose between the whole volume of water-budget and bundles of a given juice-reward volume and varying volumes of water-budget (fig. 5.11). The point of indifference between these two options gives the minimum remaining budget that must be delivered alongside the juice-reward such that the utility of that bundle is equal to the utility of the total budget. Thus, subtracting this remaining budget from the total budget should give the maximum acceptable water-budget cost, or the WTP for the juice-reward at that wealth level.

This is not analogous to the situation in the simple BC task where the monkey chooses between a given amount of budget and a given juice-reward volume. In that case, and unlike in the BDM, the monkey acquires either the juice-reward or the budget, but never both – the maximum possible offer is therefore always lower in value than in the BDM: choices were effectively made at a lower wealth-level.

The difference in wealth levels between the BC and BCb tasks, as well as the equivalence of the BDM and BCb tasks, is reflected in the mean per-trial payoff (measured in terms of water-budget) of each task (fig. 5.12). We compared payoffs of a decision-maker modelled on the monkey's behaviour in each task, as well as the payoffs of an optimal decision maker. In the BDM, the optimal decision maker always bid exactly their value for the reward, and in the BC and BCb tasks they made perfect choices with no noise – i.e. they made choices according to a step function with the step located at the decision maker's value for the reward. The BDM was simulated using a Beta distribution of (4,4) to generate computer bids – the distribution that was used in most of our BDM experiments.

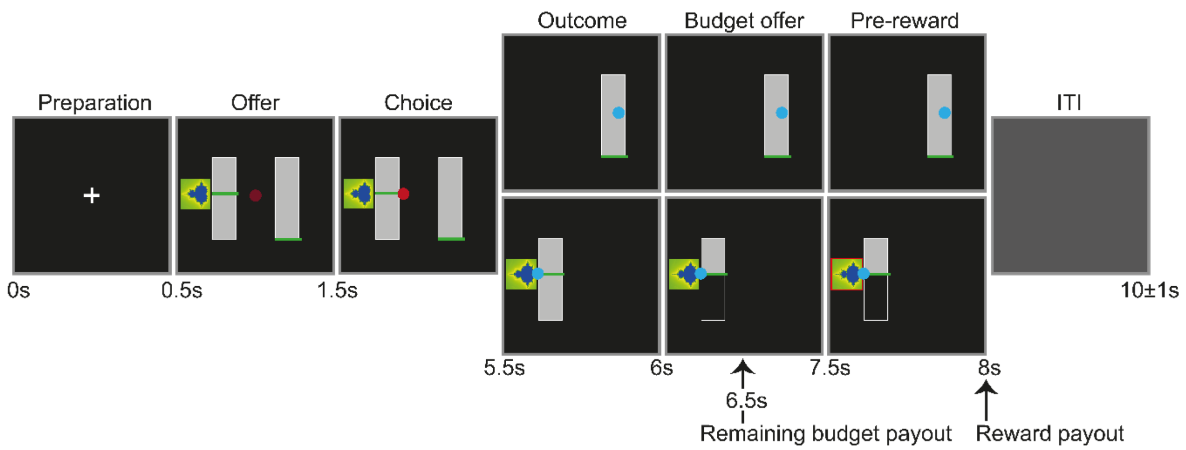


Fig. 5.11) Task design was similar to that of the simple BC task (fig. 5.1), making use of the same behavioural requirements and timings. However, in this case the water-only offer was always fixed at the total budget volume, while the bundle offer was some given reward juice and a variable volume of water-budget. The volume of water-budget in the bundle was indicated with a computer-bid marker to make the stimuli equivalent to those used in the BDM. If the bundle offer was chosen then the remaining budget would be paid out followed by the reward with delays equivalent to those used in the BDM task. Similarly, if the water-only offer was chosen then the budget would be paid out and the monkey would have to wait until the time at which reward would have been delivered before the ITI (as in the BDM when a losing bid is made). The side on which the bundle/water-only offer was presented was randomised on each trial. (Note that the stimuli appear smaller in this figure purely for presentation purposes – the stimuli were of the same size as used in previous BDM and BC tasks).

The model of the monkey's behaviour was acquired by fitting a distribution to their bids for the 0.3ml reward in the case of the BDM, or, by fitting a logistic function to the choice data in the case of the BC and BCb tasks (Appendix 4), giving us measures of decision noise in each case and a mean bid for the BDM. This allowed us to compare the BC task to the BCb and BDM tasks, as we had changed the maximum water-budget volume to 0.6ml and the juice-reward from blackcurrant to orange after calibration and testing with various combinations of budget and reward that would lead to the best use of the bidding-range*.

* A budget volume of 0.6ml of water and rewards of 0.1ml, 0.2ml and 0.3ml of orange juice were used for the BCb sessions presented in this section, as well as for the BDM sessions presented in Ch 5.4.

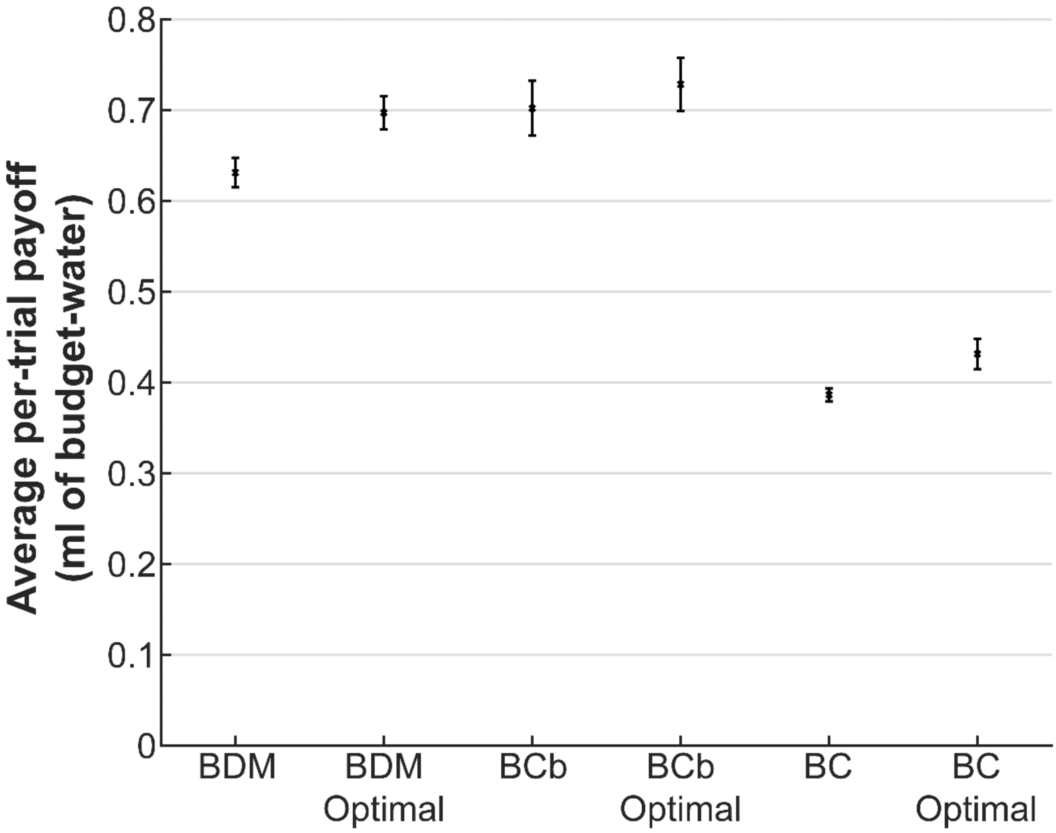


Fig. 5.12) The average per-trial payoff in terms of water-budget for a model of the monkey's behaviour acquired by fitting to samples of data for the BDM, BCb, and BC tasks as well as simulated payoffs for an optimal decision maker in each task. Average payoffs were computed for the 0.3ml reward in each task over 90 trials (the typical number of trials for the BDM/BC/BCb task in a single session when they were used concurrently), and the assumed value of that reward was taken by pooling across BCb sessions (red line in fig.5.13a).

The problem of comparing bids and choices from experiments using different volumes of budget and juice types only effects the simulations of the monkey's behaviour, which are shown for in fig. 5.12 mainly for purposes of comparison. To illustrate the difference in wealth level between the BDM/BCb and the BC tasks we only need to compare average payoffs of the simulated optimal decision-maker, for whom the average per-trial payoff was lower for the BC task regardless of the reward value that was chosen (and without requiring any assumptions regarding the variability/noise in the decision-makers choice behaviour).

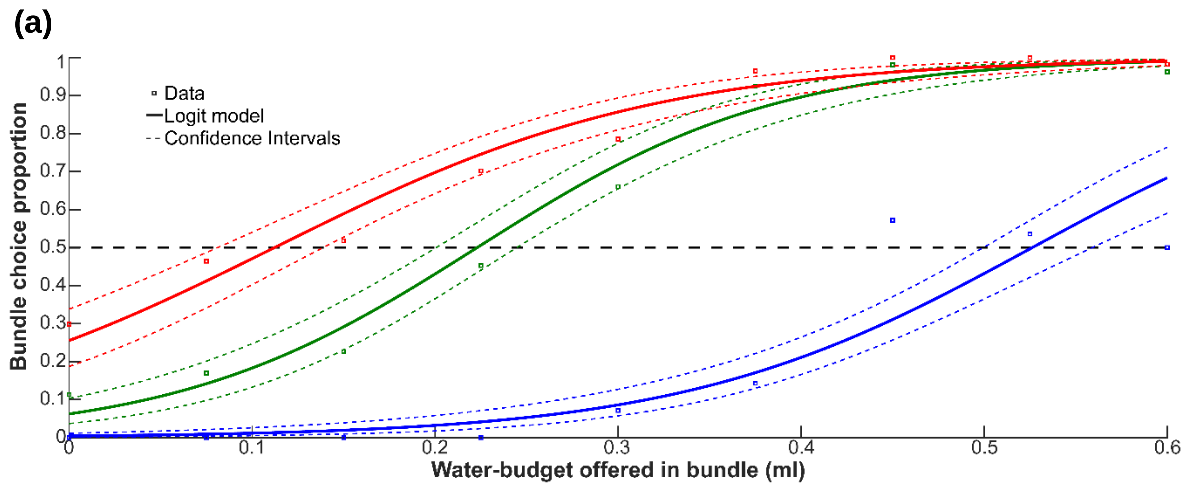
Two-sample t-tests, Bonferroni-corrected for multiple-comparisons, showed no significant difference between mean payoffs in the BDM (M = 0.69ml, SD = 0.078) and BCb (M = 0.73ml, SD = 0.14) tasks for an optimal decision maker, however, the mean payoff in the BC task was significantly lower than in either of the BDM and BCb tasks (M = 0.43ml, SD = 0.077, all $p < 0.05$). This exemplifies the problem in using values inferred from the BC task as equivalents of BDM bids, whilst showing the greater correspondence in wealth-level between the BDM and BCb tasks[*].

BCb training results:

The BCb task was run alongside an equivalent BDM task, as in the case of the simple BC task. By this point in the monkey's training the budget-volume had been changed to 0.6ml and the juice-reward was changed from blackcurrant to orange (though the same 0.1ml, 0.2ml and 0.3ml volumes were used). In particular, we hoped that the lower budget volume might allow the monkey to complete more trials per session, accelerating their learning of the task. These budget and reward parameters were used for all the sessions presented in this section and the next, which describes the results of the concurrently run BDM task, [0.6-BDM-BCb], in the first monkey, Ulysses.

We intended to compare values from the BCb and BDM tasks on a session-by-session basis, and therefore alternated between the BDM and BCb tasks in blocks of 27 trials each: 9 trials for each reward - in the case of the BCb task testing all water volumes between, and including, 0ml and 0.6ml in steps of 0.075ml. Unfortunately, the choice data were too noisy to reliably infer values of rewards on a session-by-session basis, so choices were pooled over all 14 BCb sessions for the first monkey (fig. 5.13a) and over all 20 BCb sessions for the second monkey (fig. 5.13b), and values for each reward were inferred by fitting a logistic function to these choices from across sessions.

[*] Interestingly, the monkey's simulated bidding behaviour in the BDM was significantly worse than that of the optimal decision maker but this was not the case for the BCb task, suggesting that either the degree of noise in the monkey's bidding behaviour led to significantly worse payoffs, or, that the monkey's mean bid did not match their value for the reward as inferred from the BCb task – we address the relevant BDM data in the next section.



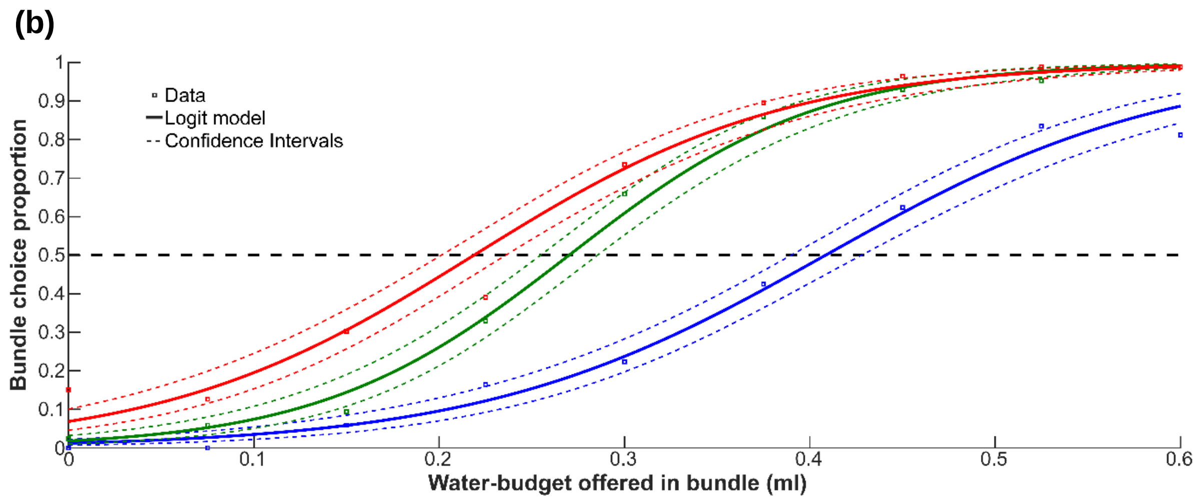


Fig. 5.13a) Choice data of the first monkey, Ulysses, pooling over 14 sessions of the BCb task. Logistic functions were fit to choices for each of the 0.3ml (red), 0.2ml (green), and 0.1ml (blue) rewards. The inferred values are equal to the total budget water volume minus the water-budget volume offered in the bundle at the point of indifference between the bundle and the full budget. This gave values of 0.49ml, 0.38ml and 0.074ml for the 0.3ml, 0.2ml and 0.1ml rewards respectively. **b)** As for (a) in the second monkey, Vicer, pooling choices over 20 sessions of the BCb task. This gave values of 0.38ml, 0.33ml, and 0.19ml for the 0.3ml, 0.2ml and 0.1ml rewards respectively.

In each case, the indifference point indicated the minimum volume of water-budget that would have to be offered alongside the juice-reward for that bundle to be equal in utility to the full budget volume. Therefore, the value of the juice-reward, and equivalent bid, was calculated by subtracting this amount from the total budget volume. This gives the maximum volume of water that the monkey is willing to sacrifice at this wealth level to gain the juice-reward, i.e. the monkey’s WTP.

Having acquired estimates of juice-reward value in terms of water-budget at a wealth level that is equivalent to that of the BDM, we were now ready to examine the monkey’s bidding behaviour in a more concrete manner, going beyond a simple analysis of the ordering of rewards by bids in terms of preference.

4 – Manipulating the computer-bid distribution

One way of encouraging a subject to approach optimality in their bidding behaviour is by using computer-bid distributions with means that are closer to the subject's value for the object being bid for – by increasing the costs of a given bid over/under the subject's value for that object (Appendix 1) - and therefore requiring reward-specific computer-bid distributions. However, we were sceptical of using such a manipulation as it could lead to bidding that approached optimality through other means. For example, if the monkey simply tracked the mean computer-bid and attempted to match their bid-marker to its location, then a computer-bid distribution with mean closer to the monkey's value for the reward would lead to bidding that was closer to optimal than for a distribution whose mean was further from this point - without the monkey having to understand or react to the payoffs in the task.

We therefore conducted some preliminary tests into the effects of different computer-bid distributions on the monkey's bids, finding that there was no evidence to suggest that the monkey was simply matching their bid to the computer's bid (Ch. 4.2). That is, the monkey's bids still differentiated between rewards of different magnitude when faced with the same computer-bid distribution for both. Therefore any 'matching' that was taking place did not outweigh the influence of reward value on the monkey's bids. This was an important finding insofar as it allowed us to manipulate computer-bid distributions without the concern that matching to the computer-bid would be the sole influence on the monkey's bids. Moreover, the results of those preliminary tests were consistent with the computer-bid distribution exerting an influence on the monkey's bids through differing expected costs of deviating from optimality – something we explore further in a subset of experimental sessions that are presented in this section.

The BDM task used here was the [0.6-BDM-BCb] task, utilising a 0.6ml water-budget and being tested alongside blocks of BCb trials. To explore the influence of the computer-bid distribution on the monkey's bids we compared a set of 10 sessions that utilised a single computer-bid distribution (SCD) for all rewards to a set of 10 sessions that used multiple computer-bid distributions (MCD), with a different computer-bid distribution for each reward-volume. Thus, we used a (4,4) Beta distribution to generate computer bids for all rewards in the SCD condition and reward specific Beta distributions of (4,2), (4,4) and (2,4) for each of the 0.3ml, 0.2ml and 0.1ml rewards respectively in the MCD condition (fig. 5.14a).

Training proceeded in alternating blocks of 5 MCD sessions followed by 5 SCD sessions to avoid any confounding effects of learning over the course of the 20 sessions. This gave a final sequence of sessions as follows: 5MCD-5SCD-5MCD-5SCD. Different conditions were not cued, so the monkey had to learn the computer-bid distribution through experience. No significant differences were found between blocks of the same type. Sessions included both BDM and BCb trials, and data from the BCb trials are presented in the previous section. The values inferred from pooling BCb trials across sessions were taken as our best estimate of the reward values in terms of water-budget for the analyses presented in this section.

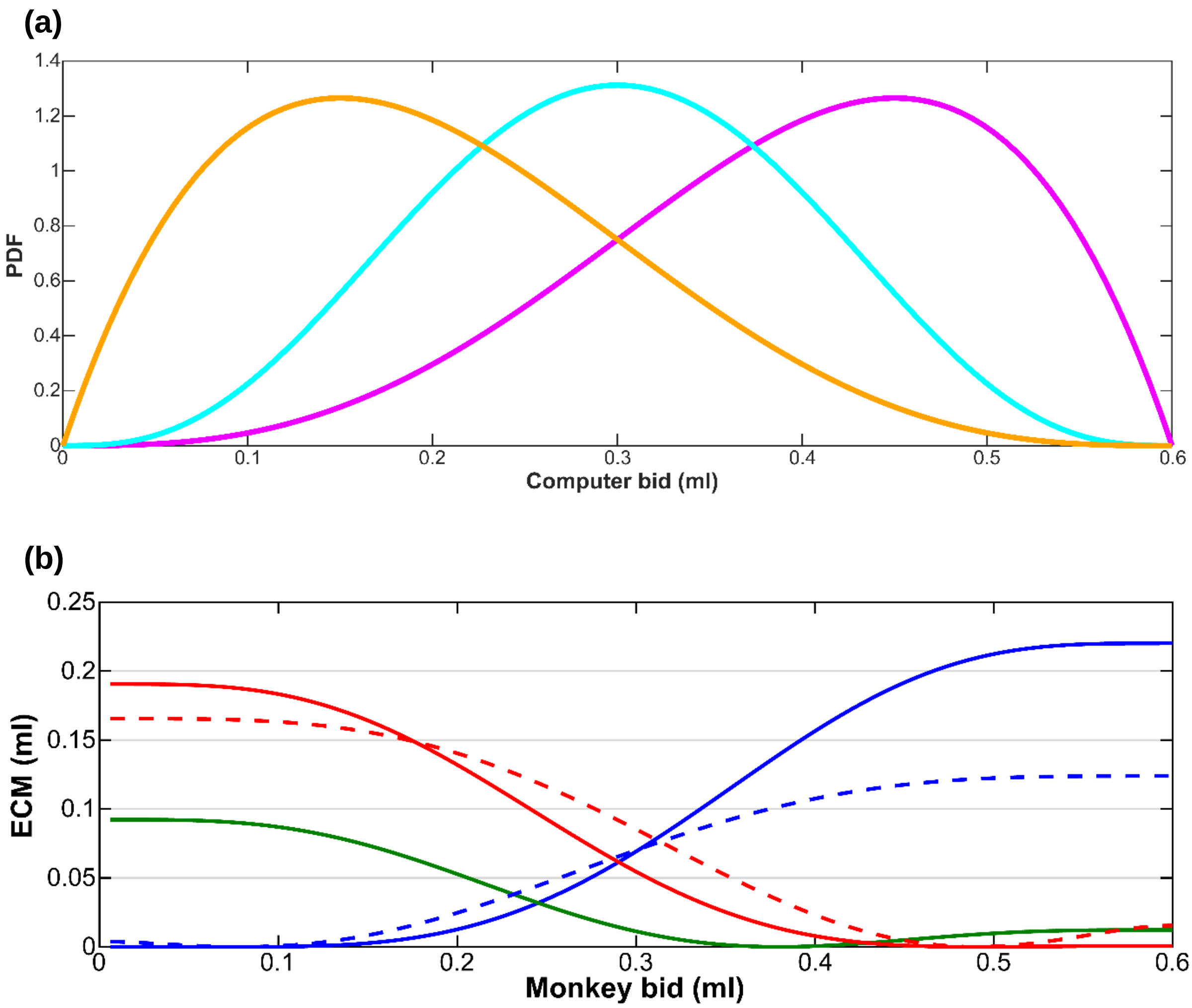


Fig 5.14a) Beta distributions used to generate computer bids for comparison of single and multiple-distribution conditions. Magenta: (4,2), Cyan: (4,4), Orange: (2,4). **b)** The expected cost of misbehaviour (ECM) for a given bid for each reward (Red: 0.3ml, Green: 0.2ml, Blue: 0.1ml) is given for both the single-distribution condition (solid lines), where the (4,4) distribution is used for all rewards, and, for the multiple-distributions condition (dashed lines), where the (4,2) distribution is used the 0.3ml reward and the (2,4) distribution is used for the 0.1ml reward – note that the (4,4) distribution is used for the 0.2ml reward in both conditions. The ECM was calculated* with the assumption that the monkey's values for each reward were the same as those inferred from the concurrently run BCb task (data shown in fig. 5.13a), such that the values of the 0.3ml, 0.2ml and 0.1ml juice-rewards in terms of water-budget were 0.49ml, 0.38ml and 0.07ml respectively.

* See Appendix 1 for ECM calculation.

Realised computer-bid distributions:

We first confirmed that the monkey was presented with computer-bids for the different rewards in the SCD condition as if they had been generated by a single probability distribution. As expected, a 1-way ANOVA found no significant effect of reward volume on the computer's bids when pooling data from across the 10 SCD sessions [$F(2,2140) = 2.18$, $p = 0.11$], and there were no significant differences in the mean computer bids (fig. 5.15a) for each reward volume (all $p < 0.05$). On the other hand, the same analysis conducted on the computer bids from across the 10 MCD sessions did find a significant effect of reward volume [$F(2,2089) = 965.62$, $p = 1.08 \times 10^{-297}$], with mean computer bids for all reward volumes (fig. 5.15b) differing significantly from one another (all $p < 0.05$).

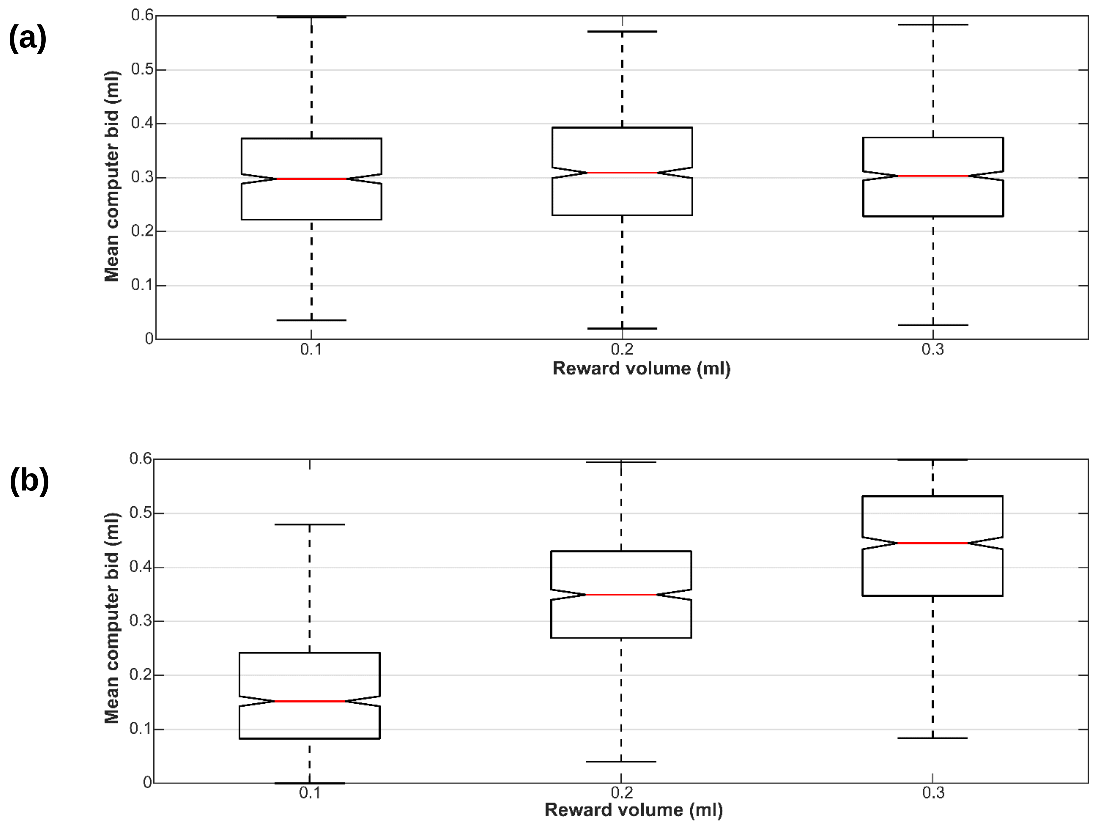


Fig. 5.15a) Mean computer bids from across all 10 sessions for each of three reward volumes in the single-distribution condition. **b)** Mean computer bids from across all 10 sessions for each of three reward volumes in the multiple-distributions condition.

<u>Sensitivity of monkey bids to different computer-bid conditions:</u>

The purpose of these experiments was to explore the influence of manipulating the computer-bid distribution, and therefore the ECM, on the monkey's bidding behaviour. Our main analyses therefore focused on measures of how well the monkey's bids reflected the reward values, and, the difference between the monkey's bids for a given reward and the optimal bid for that reward.

While ANOVA and multiple comparisons tests could be used to confirm that the monkey's bids could be used to differentiate between different reward volumes, Spearman's rank correlation between the monkey's bid and the reward volume provided a measure of the strength of this relationship and the consistency with which the monkey's bids reflected their preferences. Being able to differentiate the bids for different rewards into different clusters was something that we had already shown and did not require particularly consistent bidding behaviour, rather, it was the latter measure of consistency that we now sought to enhance in our refinement of the BDM task.

A 1-way ANOVA of bids showed a significant effect of reward volumes, thus the monkey's bids could be used to differentiate between different rewards in every session (Table 5.2). More interestingly, a comparison of the consistency of bidding in the SCD and MCD conditions showed a significant positive effect of using reward-specific computer-bid distributions, but only when pooling across sessions. A two-sample t-test found no significant difference, $p = 0.29$, between the mean Spearman's Rhos of the SCD (Mean Rho = 0.46, SD = 0.085) and MCD (Mean Rho = 0.51, SD = 0.11) conditions when the correlations are taken on a session-by-session basis.

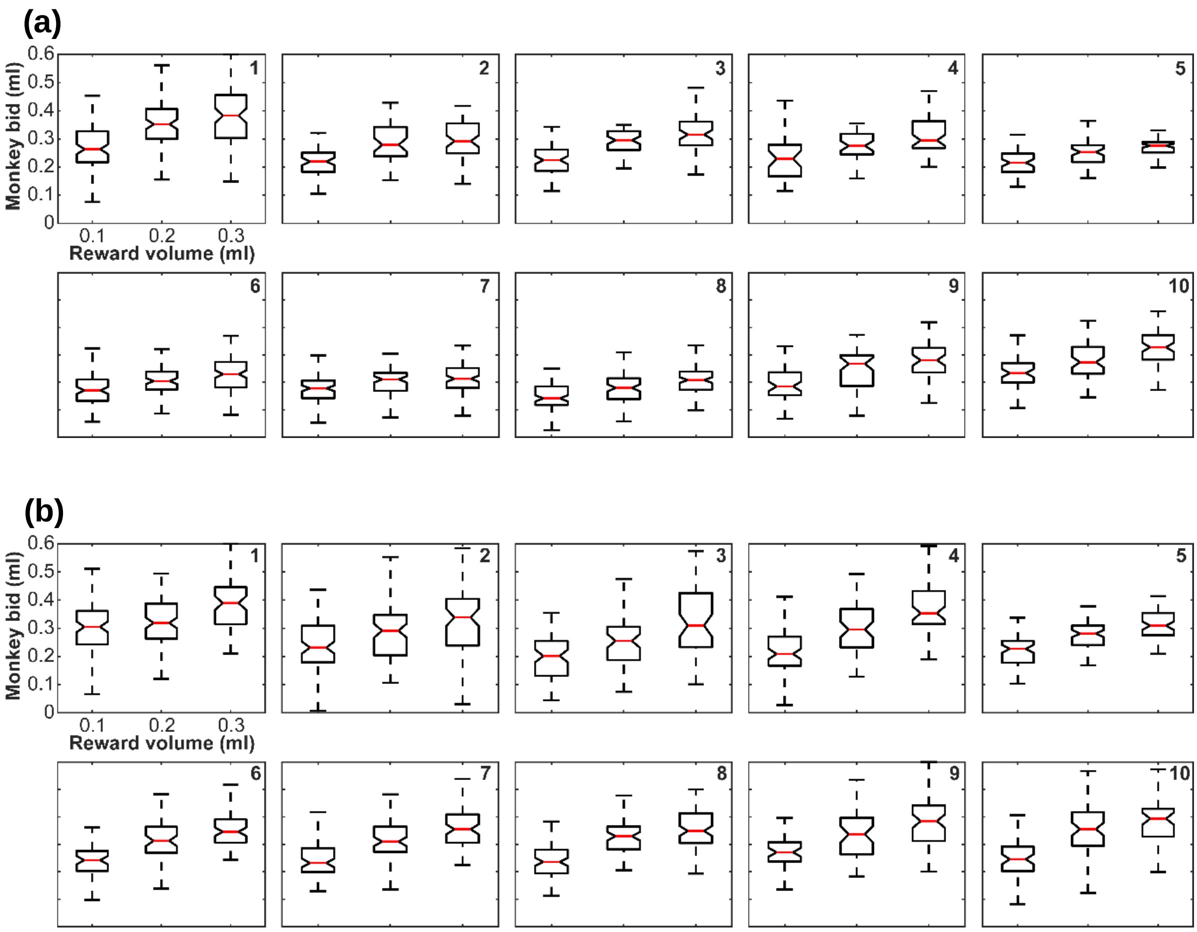


Fig. 5.16a) Mean monkey bids for each session of the single-distribution condition. **b)** Mean monkey bids for each session of the multiple-distributions condition. Note that the session number is presented in the top-right corner of each panel, with the corresponding data in Table 5.2.

Table 5.2) The mean (M) and standard deviation (SD) of bids for all sessions of both the single computer-bid distribution (SCD) and multiple computer-bid distribution (MCD) conditions. Also given are the results of a Spearman's rank correlation of the monkey's bid and the reward volume for each session (Rho and p-value), and the results of an ANOVA on the monkey's bids with factor of reward volume (both F and p values).

Session	0.3ml M (SD)	0.2ml M (SD)	0.1ml M (SD)	Rho	Spearman p	ANOVA F	ANOVA p
SCD – 1	0.37 (0.10)	0.36 (0.09)	0.27 (0.08)	0.45	1.42×10^{-14}	(2,267) = 36.72	8.17×10^{-15}
SCD – 2	0.30 (0.07)	0.29 (0.07)	0.22 (0.05)	0.50	2.33×10^{-12}	(2,169) = 30.34	5.50×10^{-12}
SCD – 3	0.32 (0.05)	0.30 (0.06)	0.23 (0.06)	0.58	1.09×10^{-18}	(2,186) = 46.59	3.96×10^{-17}
SCD – 4	0.31 (0.06)	0.27 (0.05)	0.23 (0.07)	0.49	2.01×10^{-9}	(2,132) = 22.55	3.75×10^{-9}
SCD – 5	0.27 (0.06)	0.25 (0.06)	0.22 (0.04)	0.46	3.43×10^{-11}	(2186) = 16.38	2.81×10^{-7}
SCD – 6	0.23 (0.06)	0.21 (0.05)	0.17 (0.05)	0.38	2.2×10^{-11}	(2,294) = 26.09	3.72×10^{-11}
SCD – 7	0.21 (0.05)	0.20 (0.05)	0.18 (0.06)	0.27	8.62×10^{-6}	(2,267) = 11.79	1.25×10^{-5}
SCD – 8	0.21 (0.05)	0.18 (0.05)	0.15 (0.05)	0.45	4.97×10^{-15}	(2,267) = 35.52	2.1×10^{-14}
SCD – 9	0.28 (0.06)	0.25 (0.07)	0.19 (0.06)	0.49	2.89×10^{-11}	(2,159) = 25.02	3.57×10^{-10}
SCD – 10	0.33 (0.07)	0.28 (0.06)	0.23 (0.06)	0.50	1.62×10^{-13}	(2,186) = 31.69	1.44×10^{-12}
MCD – 1	0.39 (0.09)	0.32 (0.09)	0.30 (0.11)	0.32	6.17×10^{-8}	(2,267) = 18.05	4.46×10^{-8}
MCD – 2	0.32 (0.11)	0.28 (0.10)	0.24 (0.10)	0.30	2.19×10^{-6}	(2,237) = 11.27	2.1×10^{-5}
MCD – 3	0.33 (0.12)	0.26 (0.10)	0.20 (0.09)	0.45	8.81×10^{-11}	(2,186) = 24.70	3.07×10^{-10}
MCD – 4	0.37 (0.09)	0.31 (0.10)	0.22 (0.09)	0.57	9.32×10^{-18}	(2,186) = 40.47	2.55×10^{-15}
MCD – 5	0.32 (0.06)	0.28 (0.05)	0.22 (0.06)	0.57	4.83×10^{-20}	(2,213) = 51.10	7.46×10^{-19}
MCD – 6	0.36 (0.07)	0.31 (0.07)	0.24 (0.07)	0.59	5.15×10^{-19}	(2,186) = 46.34	4.67×10^{-17}
MCD – 7	0.36 (0.07)	0.31 (0.07)	0.24 (0.06)	0.60	9.09×10^{-23}	(2,215) = 56.84	1.53×10^{-20}
MCD – 8	0.36 (0.08)	0.33 (0.06)	0.24 (0.07)	0.59	3.53×10^{-24}	(2,240) = 65.30	2.27×10^{-23}
MCD – 9	0.38 (0.09)	0.34 (0.08)	0.27 (0.06)	0.50	8.7×10^{-12}	(2,159) = 27.67	4.87×10^{-11}
MCD – 10	0.39 (0.09)	0.35 (0.10)	0.25 (0.07)	0.56	3.45×10^{-16}	(2,173) = 41.42	2×10^{-15}

Performance in the first two MCD sessions was significantly worse than that observed in the rest of the MCD sessions (Table 5.2), and it is worth noting that these were the first two sessions to be tested in this task (as 5 MCD sessions were conducted first). Outliers are often defined as those data points that are more than 3 scaled median absolute deviations (MAD) away from the median[*]. Using this criterion, session 7 of the SCD and sessions 1 and 2 of the MCD conditions appear as outliers. Excluding these sessions, one does observe a significant difference ($p = 0.01$) between the mean Spearman's Rhos of the MCD (Mean Rho = 0.55, SD = 0.052) and SCD conditions (Mean Rho = 0.48, SD = 0.056).

However, given the limited number of sessions, we sought to analyse any difference between these two conditions in a manner that would reduce the influence of individual session but without having to exclude any of them. We therefore conducted a Spearman's rank correlation on the whole population of monkey bids and reward volumes from across all sessions of the same condition. This population-level analysis showed a stronger relationship between these two variables in the MCD condition, Rho = 0.48, $p = 2.92 \times 10^{-123}$, than in the SCD condition, Rho = 0.35, $p = 8.11 \times 10^{-63}$. And this difference was confirmed to be significant using a Fisher transformation[†] of the Spearman's Rho to find the 95% confidence intervals for each of these correlation coefficients, giving bounds of [0.31, 0.39] and [0.45, 0.52] for the value of Rho in each of the single and multiple computer-bid distribution conditions respectively.

Thus, there was some reason to believe that the MCD condition did have some positive influence on the consistency of the monkey's bids, but that this effect may not strong enough to be observed on a session-by-session basis given the number of sessions that we had tested in each condition. Notably, other, later, changes to the task design (such as an increase in the water-budget volume) led to more dramatic changes in the strength of the correlation between the monkey's bids and the reward magnitudes, even when taken on a session-by-session basis.

The stronger relationship between the monkey's bids and the reward volume, observed at the population level in the MCD condition, was largely due to a greater

---

[*] See Appendix 3 for how scaled MAD is calculated.

[†] See Appendix 3 for a description of the Fisher transformation.

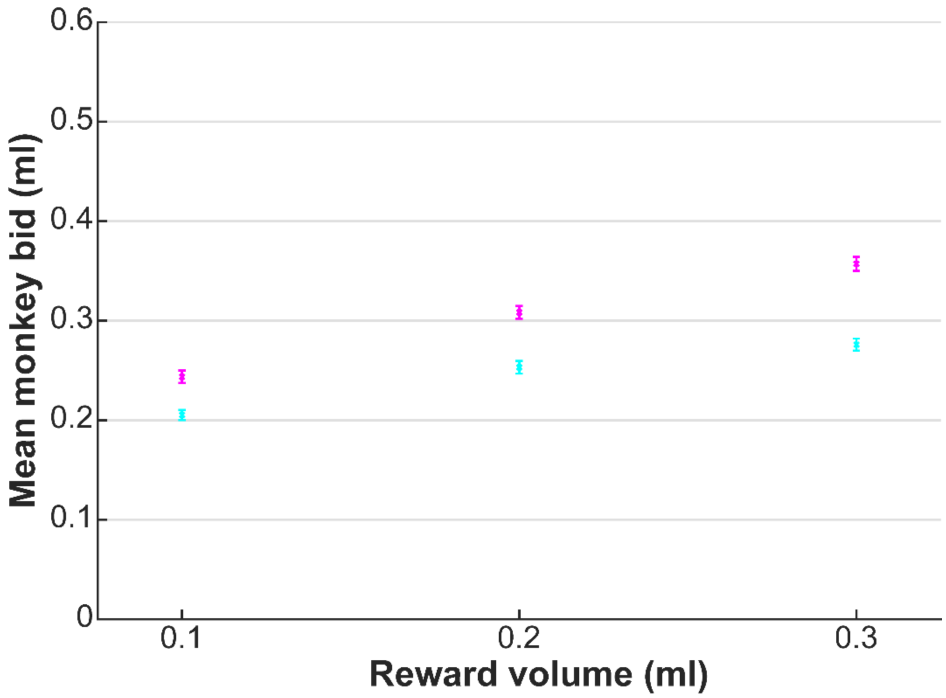
difference between the monkey's bids for different reward volumes (fig. 5.17a), as opposed to a decrease in the variance of bids for each reward volume – both of which could increase the value of Spearman's Rho by reducing the degree of overlap between the distribution of bids for each reward.

In fact, for each reward volume, there was significantly more variance in the bids in the MCD condition than in the SCD condition; two-sample t-tests comparing the mean standard deviations of bids for each reward volume from each session of each condition, found a significant difference in every case (all $p < 0.05$, fig. 5.17b and Table 5.3).

Table 5.3) The mean and standard deviations (SD) of the standard deviation and absolute bid deviance (ABD, the absolute difference between the optimal and realised bids) for each reward volume in both the single computer distribution (SCD) and multiple computer distribution (MCD) conditions. Also presented are the p values for two-sample t-tests comparing means for each volume in each condition.

Reward	SCD SD	MCD SD	p value	SCD ABD	MCD ABD	p value
0.1ml M (SD)	0.058 (0.011)	0.079 (0.018)	0.0097	0.13ml (0.034)	0.17ml (0.028)	0.022
0.2ml M (SD)	0.062 (0.012)	0.081 (0.0179)	0.012	0.13ml (0.042)	0.09ml (0.022)	0.031
0.3ml M (SD)	0.065 (0.016)	0.088 (0.0191)	0.0086	0.21ml (0.051)	0.14ml (0.024)	0.0026

(a)



(b)

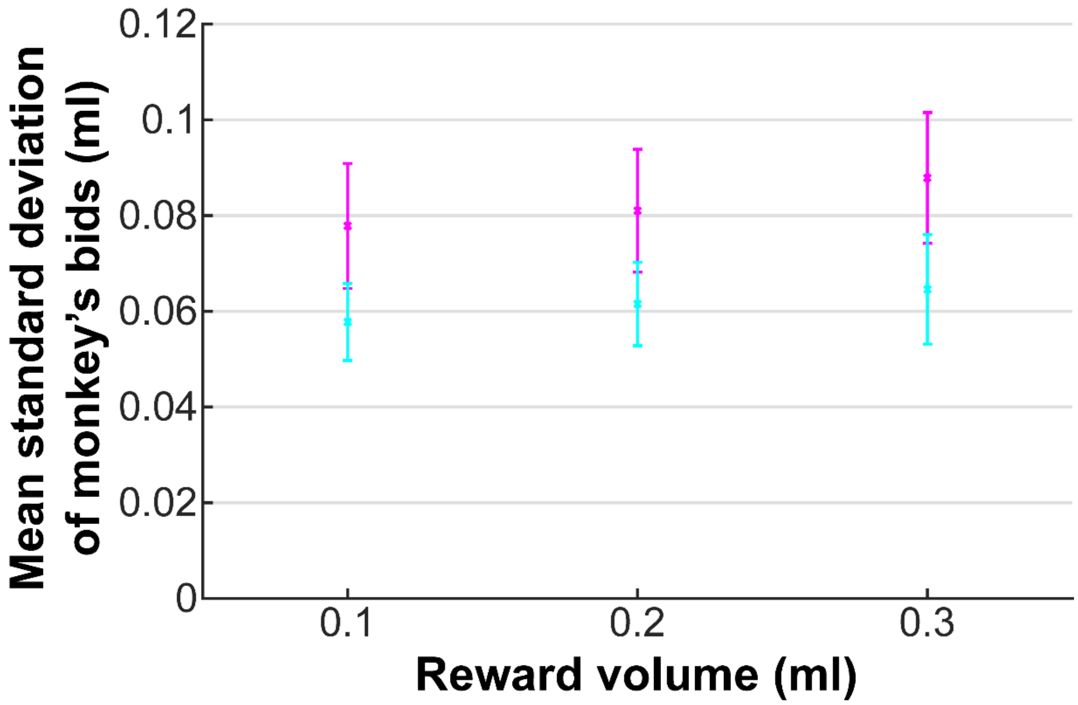


Fig. 5.17a) Mean monkey bids in the single-distribution condition (cyan) and in the multiple-distributions condition (magenta), when pooling across all sessions. **b)** The mean standard deviation of bids for each reward volume in each session is shown for both the single-distribution (cyan) and the multiple-distributions (magenta) conditions. Error bars are 95% confidence intervals of the mean. All pairwise comparisons of standard deviations for a given reward volume between the two conditions were significant.

The increased pressure to bid optimally in the multiple-distributions condition was reflected in the monkey's bids at the level of both individual sessions and when pooling across all sessions of the same condition. Overall, in the single-distribution condition, the monkey's bids were, on average, an absolute distance of 0.16ml from the optimal bid, whereas they were 0.13ml away from the optimal bid in the multiple-distributions condition; corresponding to a ~17% reduction in the distance from the optimal bid. Furthermore, different rewards were differentially effected by the use reward-specific computer-bid distributions: the mean absolute difference between the monkey's bid and the optimal bid was decreased by ~35% for the 0.3ml reward and by ~31% for the 0.2ml reward, though an increase of ~17% was observed in the case of the 0.1ml reward.

Similarly, at the session-level, the mean absolute difference between the monkey's bid and the optimal bid for a given reward also changed significantly between conditions and in a reward-specific manner (fig. 5.18). For all three reward volumes, two-sample t-tests found significant differences in this measure between the two conditions (all $p < 0.05$), without assuming equal variances of groups.

A differential effect on the tendency to approach optimal bidding is not surprising, as the payoff functions and therefore ECMs for the different reward volumes differ – more precisely, they differ significantly in the region in which the monkey placed most of his bids (46% of all bids in the single-distribution condition were between 0.2ml and 0.3ml, with 77% of all bids between 0.15ml and 0.35ml). However, this does not explain why bids for the 0.1ml reward became more distant from the optimal bid in the multiple-distributions condition, or, why a change was at all observed in the case of the 0.2ml reward – whose computer bid distribution was the same in both conditions.

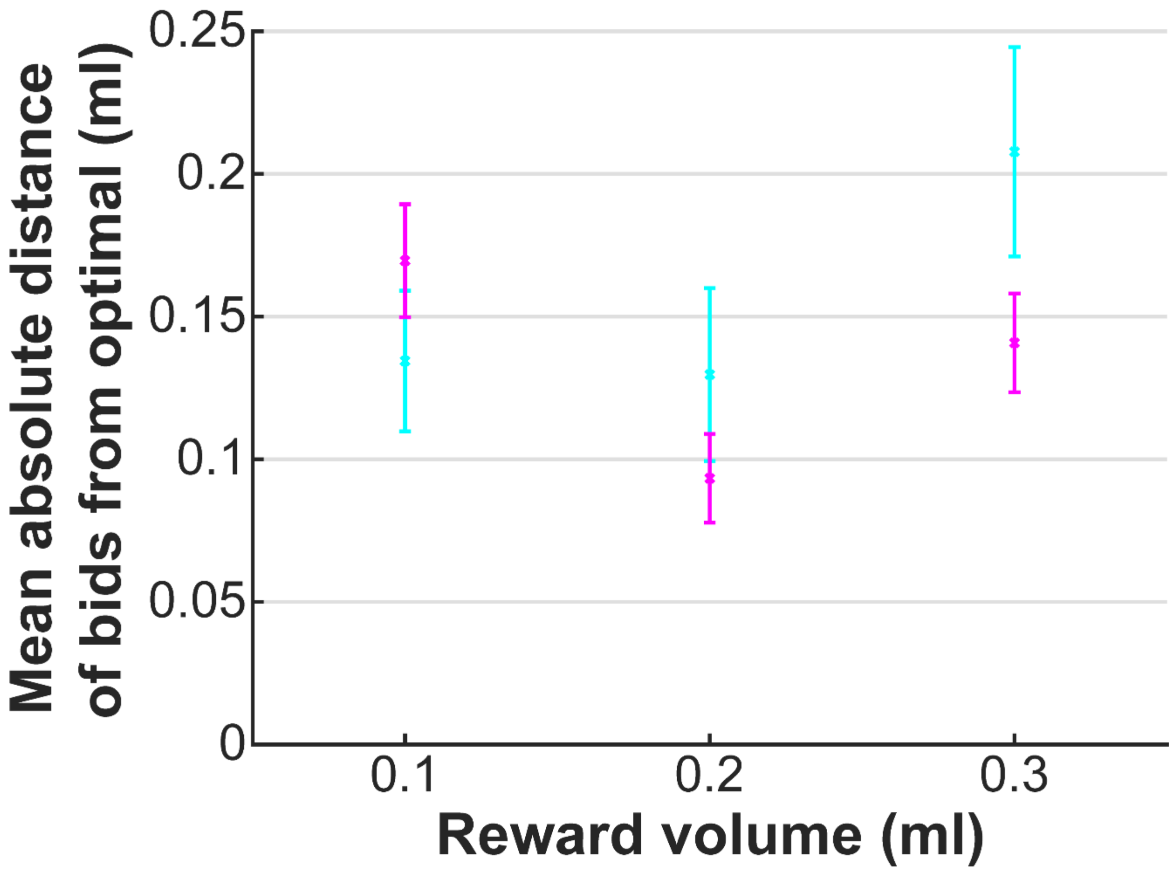


Fig 5.18) The mean absolute distance of a bid from the optimal bid for each reward volume tested in both the single-distribution (cyan) and multiple-distributions (magenta) condition. Error-bars are 95% confidence intervals of the mean.

It is possible that some generalisation across different rewards took place as the monkey came to make higher bids for all 3 reward volumes. This could explain the changes in bids for both the 0.2ml and 0.1ml rewards. Another, not mutually exclusive, explanation presents itself in the case of the 0.1ml reward: While the use of a Beta (2,4) distribution for this reward leads to a slightly higher ECM in the region of the monkey's mean bid for this reward, it also leads to a considerable reduction in the ECM for bids that are greater than 0.31ml (fig. 5.19). It is possible that reducing the punishment for over-bidding in this region outweighed the effects of the slight increase in punishment for over-bidding below 0.31ml. Moreover, this could interact with any effect of generalised higher bidding across all juice-rewards: generalised higher bidding would be incentivised for the 0.2ml and 0.3ml juice-rewards (as bids for these had been too low in the SCD condition) and would not be as strongly disincentivised for the 0.1ml reward under a Beta (2,4) distribution.

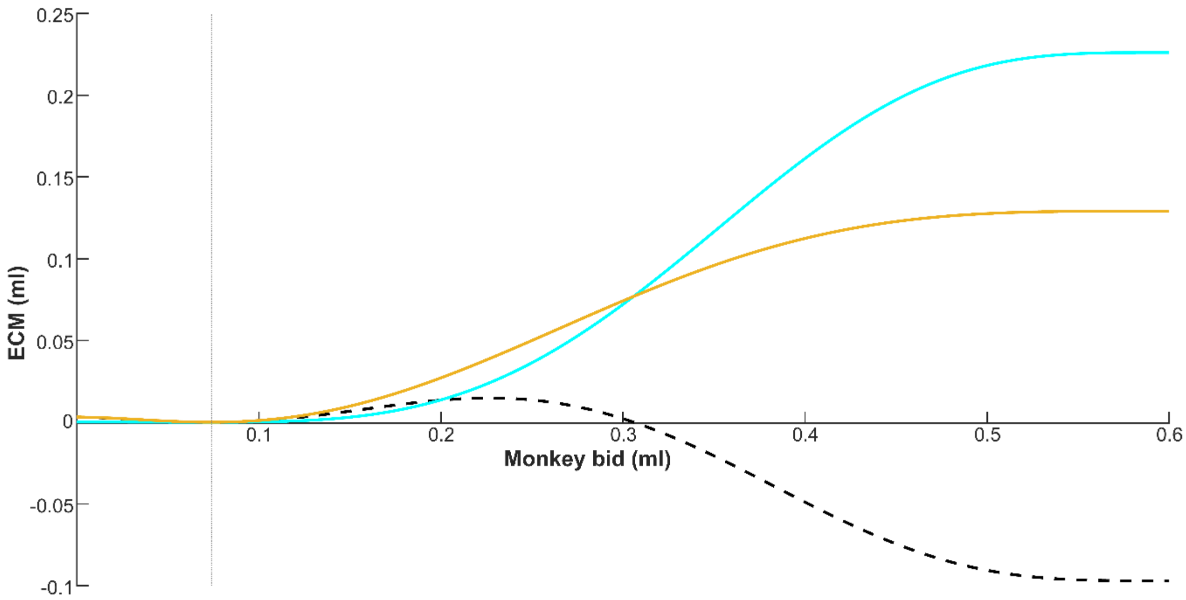


Fig 5.19) The ECMs of the Beta (4,4) distribution used in the SCD condition (cyan) and of the Beta (2,4) distribution used in the MCD condition (orange) are shown for the 0.1ml reward. The dashed black curve shows the difference between the ECMs for the (2,4) and (4,4) distributions – calculated as: ECM(2,4) – ECM(4,4). The vertical dotted line shows the point at which the ECM is zero, at the reward's water-budget value (0.074ml). The distributions whose ECMs are shown here are presented in fig. 5.14a using an equivalent colour scheme.

The monkey's pattern of bidding reflected this reduced punishment for bids higher than 0.31ml in the case of the 0.1ml reward: in the SCD condition, when faced with the Beta (4,4) distribution, the monkey only placed 7.6% of their bids for the 0.1ml reward above 0.31ml, but placed 21% of their bids for the same reward above 0.31ml when facing a Beta (2,4) distribution in the MCD condition (fig. 5.20). While this was not the intended effect of our implementation of the (2,4) distribution, these results nevertheless show a sensitivity of bidding behaviour to the payoff schedule, and suggest that the monkey may not be sensitive to the small increase in ECM of ~0.01ml that we intended to exploit in the bidding range between 0.1ml and 0.3ml.

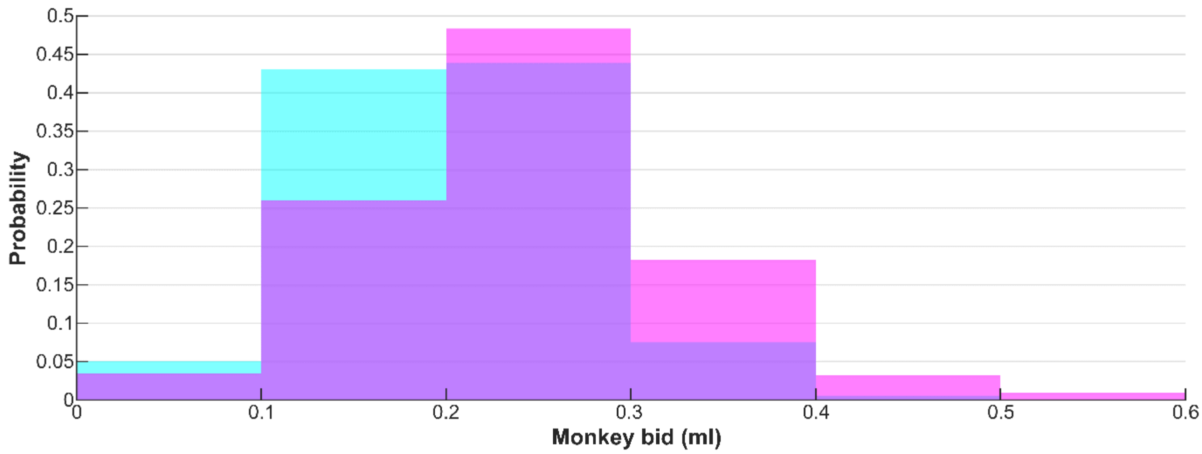


Fig 5.20) A probability histogram of the monkey's bids for the 0.1ml reward in the SCD (cyan) and MCD (magenta) conditions. Overlapping bars appear purple.

Comparison on a session-by-session basis was made difficult by the fact that sessions were not equivalent (there is no clear reason why, for example, session 2 of the MCD condition should be compared to session 2 of the SCD condition). However, pooling across sessions for each condition, bids were significantly closer to the optimal bid in the MCD condition, and this could be observed both in terms of the absolute deviation from the optimal bid, and, in terms of the proximity of the average per-trial payoff to that of an optimal bidder. The latter variable provides a useful measure of the monkey's behaviour as the monkey is likely to learn and adapt their bidding behaviour by responding to the payoffs associated with different bids (without necessarily knowing what the optimal bid is in an abstract sense).

As expected, the monkey's average per-trial payoffs more closely matched those of the optimal bidder in the MCD condition than in the SCD condition (fig. 5.21). The effect of generally under-bidding can be seen in terms of both an increased average budget-payoff on each trial (negative when subtracted from the optimal player's payoff), and a reduced average reward-payoff on each trial. And, in both cases, the monkey's payoffs are closer to that of the simulated optimal bidder in the MCD condition. The overall effect of this pattern of underbidding is the receipt of a greater per-trial volumetric intake than for the optimal bidder.

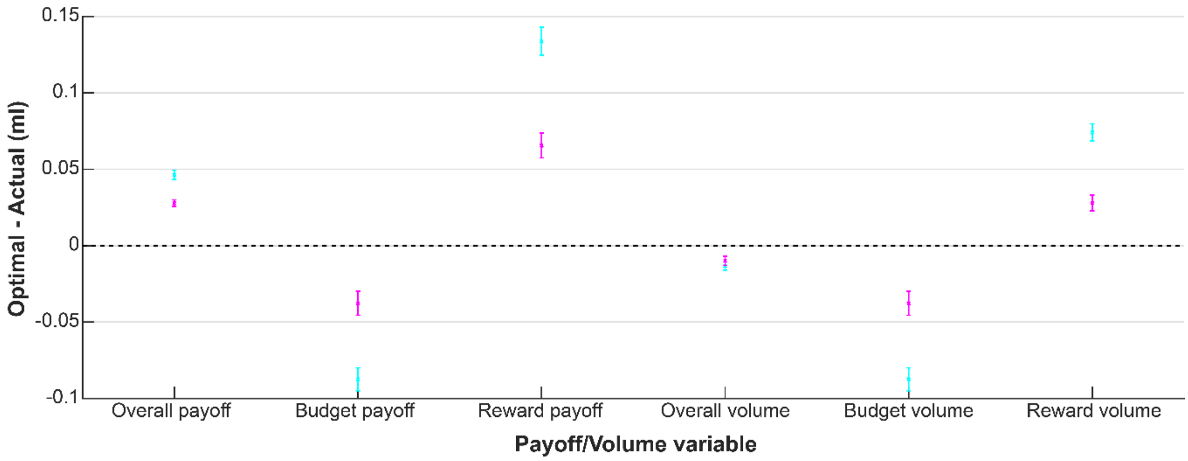


Fig. 5.21) The monkey's average per-trial payoff and volume intake, both overall and for each component in the task when compared to that of a simulated optimal bidder. Data are shown for both the SCD (cyan) and MCD (magenta) conditions. Error-bars are 95% confidence intervals of the mean.

Importantly, the difference between the overall payoff in the MCD condition and the payoff of an optimal bidder is only ~0.03ml per trial, and this is comparable to the standard deviation of volume delivery in our solenoid juice/water delivery systems (Appendix 2). It is therefore possible that the manipulation of computer bids can only have a limited effect in driving more accurate bidding at this point, as the monkey may not be able to distinguish a difference between the payoff when bidding optimally as opposed to bidding as they were in the MCD condition. Indeed, such a small ECM may even be smaller than the difference that can be perceptually discriminated by the monkey.

To complicate matters, such a discrimination must be made on the basis of an expectation of outcomes over many trials, and not simply on a trial-by-trial basis. Take for example the bids for the 0.3ml reward in the MCD condition. Even pooling across all 672 trials, the mean payoff for an optimal bidder, in units of water-budget, is 0.68ml with a standard deviation of 0.098ml, whilst the mean payoff of the monkey for this reward is 0.66ml, with a standard deviation of 0.1ml. Such close mean payoffs are observed despite the fact that on average the mean absolute deviation of the monkey's bids from the optimal bid for this reward is 0.14ml (Table 5.3), ~25% of the total bidding range!

Taken together, these results show a relatively weak but significant effect of manipulating computer-bid distributions on the monkey's bids. A greater number of sessions and a more systematic study of these effects is required to draw more extensive conclusions – and this presents a potentially fruitful avenue for future study into the effects of costs on bidding behaviour (Ch. 6.1). However, for the purposes of refining the BDM task to improve bidding accuracy the current data suggest a limited utility of manipulating the computer-bid distributions – especially considering the relatively low expected costs of misbehaviour (ECM), even for relatively large deviations from the optimal bid (Appendix 1).

5 –Top-start (TS) position BDM

By this stage of training of the BDM task we had established that the monkey's mean bids for different reward volumes could be used to correctly rank those rewards in order of preference. However, there was an important confound that had to be ruled out before we made further changes to improve the consistency of the monkey's bids: With the starting position of the bid-marker at the bottom of the budget-bar, the distance that the marker had to move to correctly rank rewards by their bid was directly proportional to the value of that reward. Thus, the monkey might be able to rank rewards correctly in the BDM simply by pushing the joystick with a force that was proportional to the motivational value of the reward. And, as it has been previously shown that response vigour is related to the reward value for which the operant behaviour is produced[57,58], this a potentially significant confound. If the monkey was simply pushing the joystick harder when presented with a higher value reward, then we might observe rank-ordered bids without the monkey having any understanding of the task or the payoffs implied by various bids.

The results of our previous analysis of computer-bid manipulations (Ch. 5.4) were consistent with the monkey's bids being sensitive to the payoffs in the task and did suggest some understanding of the task contingencies. However, to confirm that rank-ordered bids were not simply a by-product of response vigour, we would have to decorrelate response vigour and the monkey's bids. We achieved this by reversing the starting position of the monkey's bid-marker, such that bidding started at the top of the budget-bar, [TS-0.6-BDM-BCb].

In this case, the highest bids would require the smallest movements of the bid-marker, and therefore of the joystick, whilst the lowest bids would require the largest displacements of the bid-marker. If the monkey was simply moving the joystick with a vigour of response proportional to the reward value, then we would now expect the highest bids to be made for the lowest value reward and vice versa.

It was also possible that the monkey had learnt a bid-marker position for each juice-reward fractal, in which case they might simply reproduce this final bid-marker position from another starting position. We therefore trained the monkey with new fractal images that had only been seen in 3 Pavlovian stimulus learning sessions

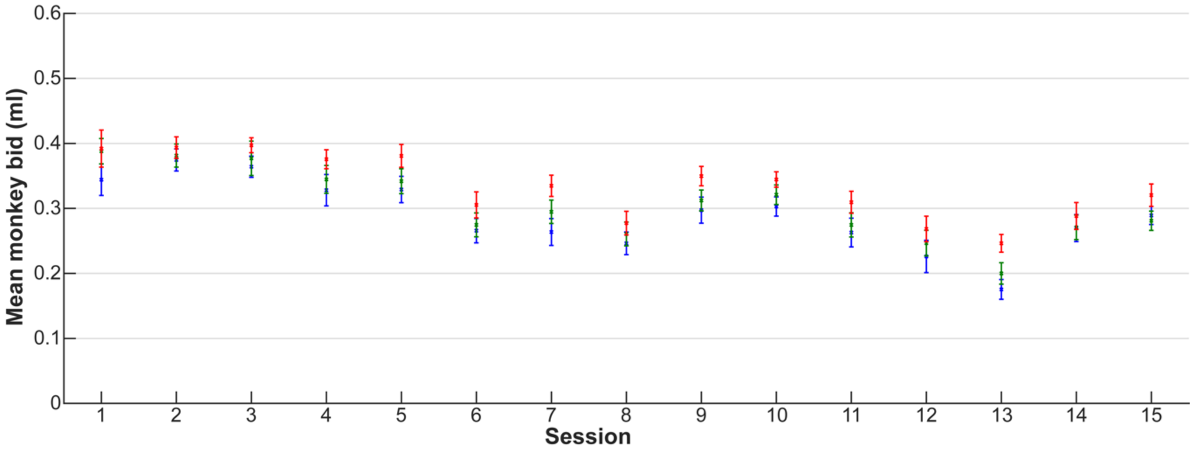


Fig. 5.22) The mean bids for each session in the [TS-0.6-BDM-BCb] task in the first monkey, Ulysses. Mean bids are shown for the 0.3ml (red), 0.2ml (green), and 0.1ml (blue) rewards. Error-bars are 95% confidence intervals of the mean.

preceding the TS BDM training (but never in a BDM task). The monkey had also been trained in moving the bid-marker from the top of the budget-bar in a preceding target-task, but different juice-reward fractals were used in those sessions. Thus, any correct ranking of the rewards could more reliably be attributed to an understanding of the task structure and payoffs, rather than the simple reproduction of a learnt stimulus-response relationship.

The results of the [TS-0.6-BDM-BCb] task for the first monkey were consistent with bidding resulting from some understanding of the task structure and payoffs, rather than being a simple by-product of some relationship between reward value and the vigour of response, and this was apparent from the first session that utilised the new starting position (fig. 5.22), despite the use of new reward-predicting stimuli.

However, this was only true in the limited sense that the overall ordering of bids was rational in most sessions, with 1-way ANOVAs finding a significant effect of reward volume on bids in all but 3 sessions (sessions 2, 8, and 14), but a multiple comparisons test could not reliably differentiate bids for all 3 rewards in any single session. Moreover, the value of Spearman's Rho for the correlation of the monkey's bid and the reward volume was low in individual sessions (with a minimum value of 0.07 in session 14, and a maximum value of 0.39 in session 13), giving a mean Spearman's Rho of 0.23 (SD = 0.087) across sessions.

While the bidding performance was clearly significantly worse than in previous versions of the BDM, the data nevertheless suggested that the monkey immediately knew to bid more for the highest value reward than the lowest value one, despite the now inverted relationship between effort expended and the magnitude of the monkey's bid.

Therefore, to more concretely establish whether the vigour of response accounted for the patterns of bidding observed in the BDM, we would need to seek a general improvement in the monkey's ability to differentiate their bids in the task - noting that bids in the equivalent bottom-start version of the task preceding this were also poorly differentiated. With regards to this it is worth noting that the mean ECM of the monkey's bids in this task (M = 0.047ml, SD = 0.082ml) were not significantly different to those from the single computer-bid distribution condition (M = 0.046ml, SD = 0.073ml) described in the previous section. If the monkey was not sensitive to this magnitude of costs, then we may not expect any improvement in their bidding behaviour beyond this level and would have to seek other means of improving their bidding precision before exploring the effects of starting position, and the possible confound of movement vigour, on bidding behaviour[*].

[*] We delayed the training of a TS BDM task in the second monkey until after we had improved the consistency and accuracy of bidding behaviour by the means described in section 6 of this chapter – where we describe the results of a later version of the TS BDM task using new parameters.

6 – Changes leading to the final version of the BDM

In this section, we describe further iterations of the task that made use of different parameters and stimulus aids and saw a dramatic improvement in the monkeys' bidding performance. Due to time constraints, we had to change several parameters at once, as we were unable to implement each change alone followed by a sufficiently long period of testing. This makes it difficult to attribute any improvement in performance to a single parameter change or manipulation of the task structure. Moreover, we could not account for the effect of increased experience with the BDM task when performing these latest versions of the task relative to when they were first trained in the version of the task presented in Chapter 4.

After establishing that the BDM could be used to differentiate rewards in terms of their order of preference (Ch. 4.2), we tested several different versions of the task, presented in the preceding sections of this chapter. Those experiments suggested that a continuous bidding paradigm utilising a joystick effector (as in our preliminary tests of the BDM) produced the most consistent and well differentiated bids. Moreover, a cursory analysis of the effects of using reward-specific computer-bid distributions showed a very limited positive effect on the consistency of the monkey's bids and their proximity to the monkey's value for the reward as inferred from the binary-choice-bundle (BCb) task. Therefore, it was likely that improvements in the monkeys' performance would require other manipulations of the task parameters.

Notably, performance of the first monkey in the preliminary version of the task was significantly superior to that observed in any of the alternative versions of the task hitherto described in this section - as measured by the strength of the Spearman's rank correlation between the monkey's bids and the reward volume. There are two key interrelated changes between that preliminary version of the task and those that would follow that could explain this: a decrease in the volume of water-budget on each trial, and, concurrent performance of the BDM and BCb tasks.

The total available budget with which bids could be made was reduced for three reasons. First, our preliminary tests using both high and low budget-volume versions of the task (Ch. 4.2) showed a dramatic improvement in the case of the latter, lower volume, BDM task. We speculated that this could be due to the increased number of trials that the monkey could complete in each session before becoming sated, thus granting the animal more experience of the task contingencies. Second, as we ultimately aimed to produce a task that could be used in a neuronal recording study, we were incentivised to reduce the payouts on any given trial so that we could maximise the number of trials within which an isolated neuron could be tested. Third, the introduction of concurrent BCb testing necessitated a reduction in the per-trial payouts if we were to be able to collect enough data from each of the two methods in a single day of testing.

By simultaneously testing both the BDM and BCb tasks we had hoped to establish session-by-session estimates of the monkeys' value for each reward from each method. In theory, this would allow for a more principled comparison of the values inferred from each method, which, for example, should be equally affected by day-to-day fluctuations in the monkeys' values (e.g. due to changing levels of satiety). However, in practice, this may have been too ambitious, or, introduced too early in the monkeys' training – indeed, the requirement of flexibly switching between different tasks can reduce performance in each. Ultimately, the theoretical advantages of concurrent BDM and BCb testing cannot be realised if the overall consistency and accuracy of the monkeys' bidding and choice behaviour is too poor to lead to reliable differentiation of rewards.

Therefore, we now chose to test the BDM and BCb tasks on separate days, using a block of BCb sessions before and after the BDM sessions and inferring reward values from the choices made in the BCb sessions across both blocks, providing a BCb value estimate that accounts for changes in the monkeys' values for the rewards over the BDM testing period.

In turn, we could now also use larger volumes of water-budget on each trial. Although this has no effect on the relative expected costs of misbehaviour for a given deviation from the optimal bid, the absolute costs of a bid made with the same relative distance from optimal is higher. For example, consider a bidder with a value of 0.5ml bidding between 0 and 1ml against a computer using a Beta (4,4) distribution to generate its bids. In that case a bid of 0.6ml has an expected cost of 0.012ml, or 1.2% of the total budget. The equivalent relative distance when the budget is doubled to 2ml is a bid of 1.2ml for a reward with a value of 1ml, in this case the cost is 0.024ml which is also 1.2% of the total budget.

Increasing the budget volume could therefore lead to greater absolute costs for the same relative deviations from the optimal bid, as monkeys' preferences are influenced by absolute volumes of reward, not just relative gains/losses. This greater absolute difference in reward value may increase the learning rate, driving bids towards the optimum. In fact, this effect could be more/less pronounced than would be predicted from an observation of the difference in absolute volume alone if the utility gradient between the two bids/actions drives the rate at which the probability of choosing one action comes to dominate another[*].

While the first monkey, Ulysses, was being trained in the multiple computer-bid distribution (MCD, section 4) and top-start versions of the BDM (TS BDM, described in section 5 of this chapter), the second monkey, Vicer, was trained in a version of the task that utilised a larger water-budget volume of 0.9ml. Those sessions showed an improvement in the monkey's bidding relative to a previous version of the task using a 0.6ml water-budget and concurrent BCb testing.

Following this, both monkeys were introduced to a newer version of the BDM task with an even larger water-budget volume of 1.2ml, as well as several other changes to the task timings and display that we speculated might improve the monkeys' performance.

Throughout this section, we define the strength of the monkey's 'performance' in the BDM as the value of Spearman's Rho for the correlation between the monkey's bids and the reward volumes, and report the mean Rho and its standard deviation (SD) across all sessions of the same type. When comparing different versions of the BDM task, we use 2-sample t-tests of the two sets of Spearman's Rhos, without assuming equal variances. All sessions described in this section used a Beta (4,4) distribution to generate computer bids.

[*] This is discussed in greater detail in Appendix 1 where the concept of ECM introduced by Lusk et al. (2007) is extended to account for non-linear utility of the budget, with a discussion of the implications of this for the rate of learning in the BDM task.

Increasing budget-volume improved performance in Vicer:

The second monkey, Vicer, had previously been trained in the version of the BDM that was used for the single computer distribution (SCD) condition in Ulysses. That BDM task was tested alongside an equivalent BCb task in the same sessions, utilising 0.6ml of water-budget; 0.3ml, 0.2ml, and 0.1ml blackcurrant juice-rewards, and, a continuous bidding paradigm [0.6-BDM-BCb]. This was Vicer's first set of sessions using a continuous version of the BDM, as he had first been trained in a discrete version of the task (Ch. 5.2).

A significant Spearman's rank correlation between the monkey's bids and the reward volumes was found in only 1 of the 10 [0.6-BDM-BCb] sessions, with Rho = 0.25 ($p = 0.011$), and, across sessions, mean Rho = 0.1 (SD = 0.062).

Vicer was then tested in 10 sessions [0.9-BDM-BCb] with an increased water-budget volume of 0.9ml, also alongside an equivalent BCb task (fig. 5.23). The blackcurrant juice-reward concentration was increased from 1:9 (concentrate:water) to 2:9, to increase the relative value of the 0.3ml, 0.2ml, and 0.1ml juice-rewards, making use of the wider budget range. The Spearman's rank correlation between the monkey's bids and the reward volumes was significant in the last 6 of these 10 sessions, and had a mean value Rho = 0.31 (SD = 0.26). 2-sample t-tests found that this was significantly greater than the mean Rho for the preceding 0.6ml version of the task ($p = 0.034$), however, it was not significantly greater than that of either the 2-reward (fig. 5.8) or 3-reward (fig. 5.9) discrete versions of the task which also used a 0.9ml water-budget ($p = 0.86$, and $p = 0.062$ respectively).

It is worth noting that a sharp improvement in performance was observed after session 4 of the [0.9-BDM-BCb] task, and, if the first 4 sessions are excluded - giving mean Rho = 0.49 (SD = 0.10) then performance is found to be significantly greater ($p = 0.0024$) than that observed in 6 sessions of the discrete BDM task for which there was also a significant relationship between bids and reward volumes[*], mean Rho = 0.28 (SD = 0.065).

[*] Here we pool all 5 of the 2-reward discrete-BDM sessions and one 3-reward discrete-BDM session.

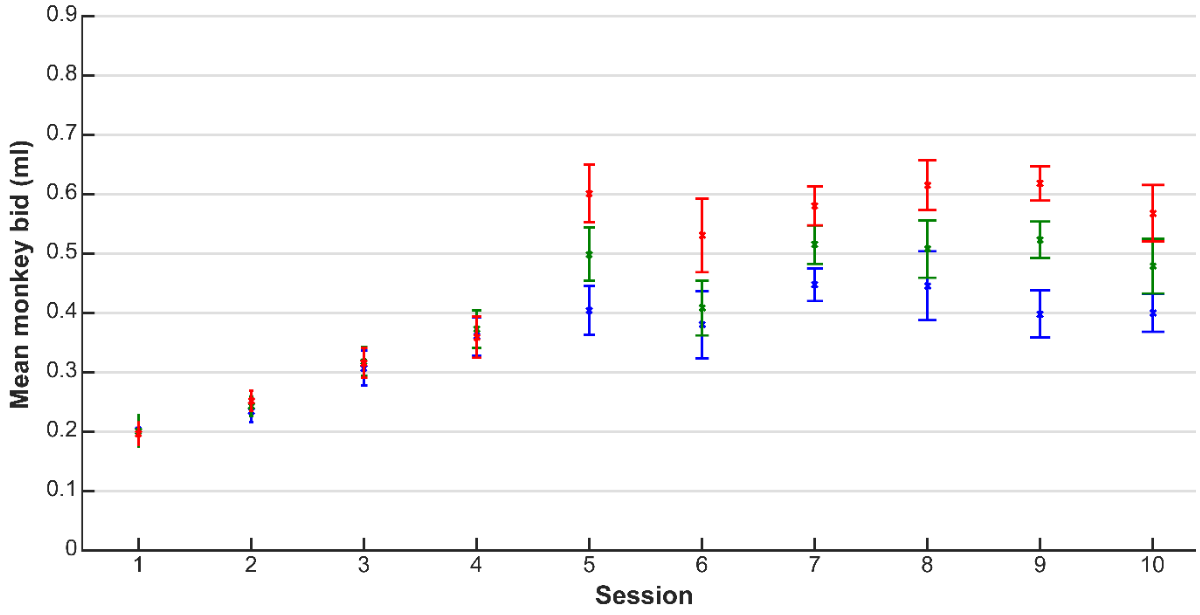


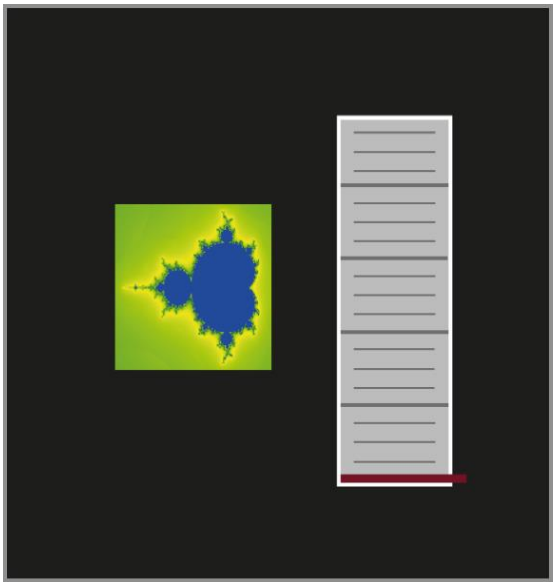
Fig. 5.23) Bids from across 10 sessions of the [0.9-BDM-BCb] task with a water-budget volume of 0.9ml and blackcurrant juice-rewards of 0.3ml (red), 0.2ml (green), and 0.1ml (blue). Error-bars are 95% confidence intervals of the mean.

These results only showed that Vicer's bidding had improved since he had been trained in the discrete BDM task, but could not determine whether this was due to a general improvement in performance over time, or, due to the use of a continuous, rather than discrete, bidding paradigm.

Nevertheless, further improvements in the monkey's performance in the continuous version of the task were observed when we began training the BDM alone, without concurrent blocks of the BCb task and with an even larger water-budget volume of 1.2ml [1.2-BDM-A]. As such, the rewards again had to be adjusted to make use of the whole bidding range – as assessed by the results of BCb testing. We found that mango flavoured rewards of 1ml, 0.5ml and 0.15ml had values that covered a broad range of the 1.2ml water-budget. From a practical point of view, using a higher water-budget volume also effectively allows for easier selection of juice-reward volumes whose values will cover the budget-range, as increasing the volume difference between different rewards reduces the relative effects of noise in our solenoid delivery system (a limiting factor for the minimum discriminable differences in volume that can be delivered – see Appendix 2).

For these sessions, we also introduced scale-lines on the budget-bar (fig. 5.24a), to assist the monkey in recognising and learning profitable relative positions of the bid-markers.

(a)



(b)

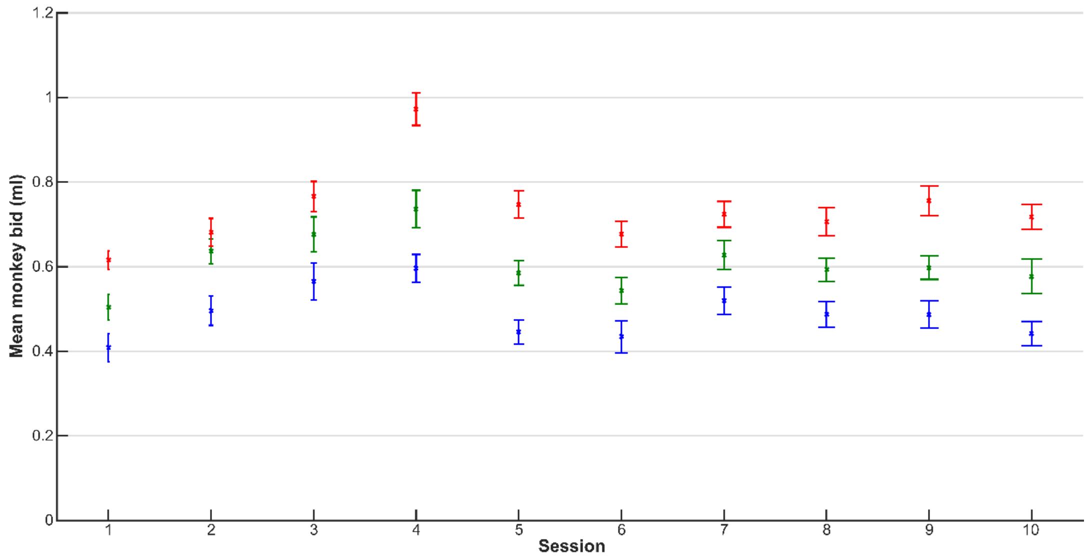


Fig. 5.24a) A schematic of the ‘offer’ epoch of the BDM task, showing the addition of scale-lines onto the budget-bar stimulus for the [1.2-BDM-A] task. **b)** Mean bids from 10 sessions of the [1.2-BDM-A] task, which introduced scale-lines and yet another increase in the water-budget volume, to 1.2ml. Mean bids are shown for the 1ml (red), 0.5ml (green) and 0.15ml (blue) mango juice-rewards, error-bars are 95% confidence intervals of the mean.

The Spearman's rank correlation between the monkey's bids and the reward volume was significant for these sessions (all $p < 0.01$), with mean Rho = 0.65 (SD = 0.09). Performance across these 10 [1.2-BDM-A] sessions (fig. 5.24b) was significantly superior to that of the preceding [0.9-BDM-BCb] task ($p = 0.002$), even if the first 4 [0.9-BDM-BCb] sessions were excluded ($p = 0.014$).

Finally, Vicer was trained in a version of the task that was identical to the [1.2-BDM-A] task but with bids starting from the top of the budget-bar, [1.2-TS-BDM-A] (see section 5). Performance in these 10 sessions (mean Rho = 0.47, SD = 0.28) was not significantly worse than with the preceding bottom bid-marker starting position (2-sample t-test $p = 0.085$), and was notably better than that observed in the first monkey, Ulysses, when he was tested in a lower water-budget volume version of the top-start BDM (mean Rho = 0.23, SD = 0.087, two-sample t-test $p = 0.025$ – see fig 5.22). Bids reflected the monkey's order of preference for the rewards in the last 8 sessions of this task (fig. 5.25), with Bonferroni-corrected multiple-comparisons tests showing a significant difference in bids for all groups in session 4 as well as in sessions 6-10.

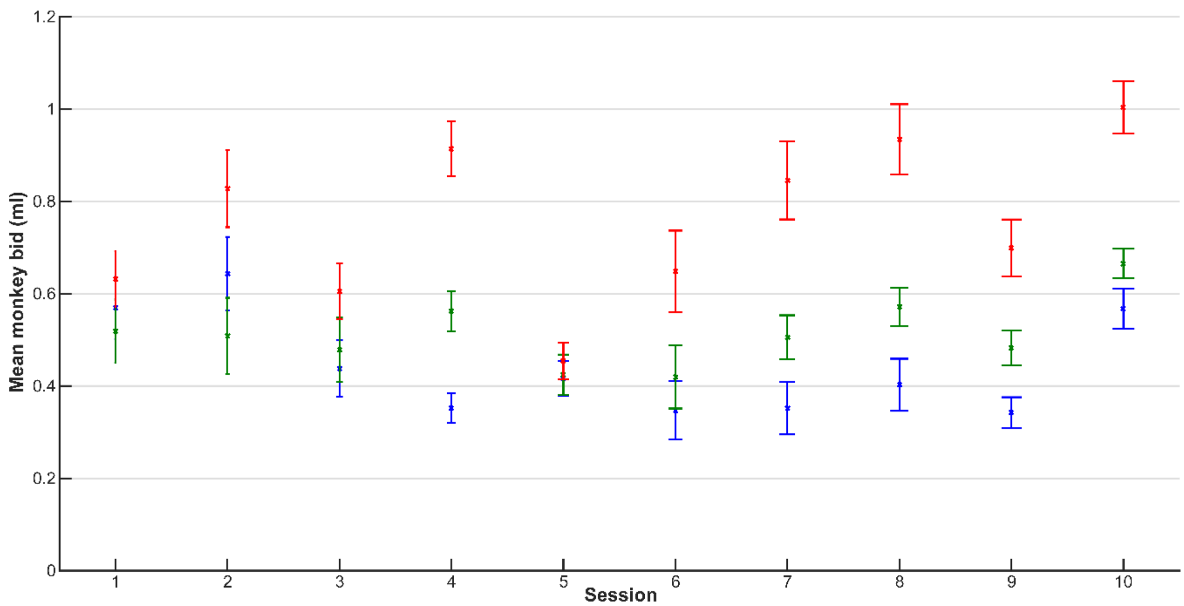


Fig. 5.25) Mean bids from 10 sessions of the [1.2-TS-BDM-A] task. Mean bids are shown for the 1ml (red), 0.5ml (green) and 0.15ml (blue) mango juice-rewards, error-bars are 95% confidence intervals of the mean.

Further improvements were observed with task-timing changes:

Having observed an improvement in both bottom-start and top-start versions of the BDM task in Vicer, we now decided to train both monkeys in a version of the task using a higher budget-volume and without concurrent BCb training. However, in addition to the changes already implemented and tested in Vicer using the [1.2-BDM-A] task, we now introduced a randomly generated delay, or ‘jitter’, in the timing of the ‘offer’ epoch, increased the presentation time of the computer-bid from 0.5s to 1s, and, reduced the inter-reward interval (IRI)* from 1.5s to 0.5s, [1.2-BDM-B] (fig. 5.26a). Equivalent timing changes were also applied to the BCb task.

The ‘jitter’ was generated from an exponentially distributed variable delay, making it difficult for the monkeys to predict the exact timing of when they could move the joystick. This should prevent the monkeys from preparing a response in the ‘offer’ epoch and then moving the bid-marker without attending to the screen at the known time of the ‘bidding’ epoch, a behaviour we sometimes observed them, presumably lazily, engaging in. Attempting to do this in the presence of such a ‘jitter’ would lead to more errors and a reduction in the reward-rate as the joystick would then often be moved during ‘offer’ epoch, leading to a ‘not-centred’ error (fig. 5.1).

In previous versions of the task (e.g. fig. 5.26b) we observed that the monkeys would often stop attending to the screen and reorient themselves to focus on the reward and budget-delivery spouts soon after bidding had ended. There was only a 1s delay between the end of the choice epoch and delivery of the remaining budget, and we were concerned that this was stopping the monkeys from attending to the outcomes of their bids – diminishing their learning of the task contingencies. Therefore, to emphasise the relationship between the relative positions of the bid-markers and the win/loss contingencies of the task, we also increased the length of the ‘computer bid’ epoch, from 0.5s to 1s. Furthermore, the remaining budget was now paid out at the end of the ‘budget offer’ epoch, after the paid out portion had been occluded for 1.5s. Overall this increased the total time from the end of the ‘choice’ epoch to the time of receiving the remaining budget from 1s to 2.5s.

* The IRI is the delay between receipt of the remaining water-budget and receipt of the juice-reward.

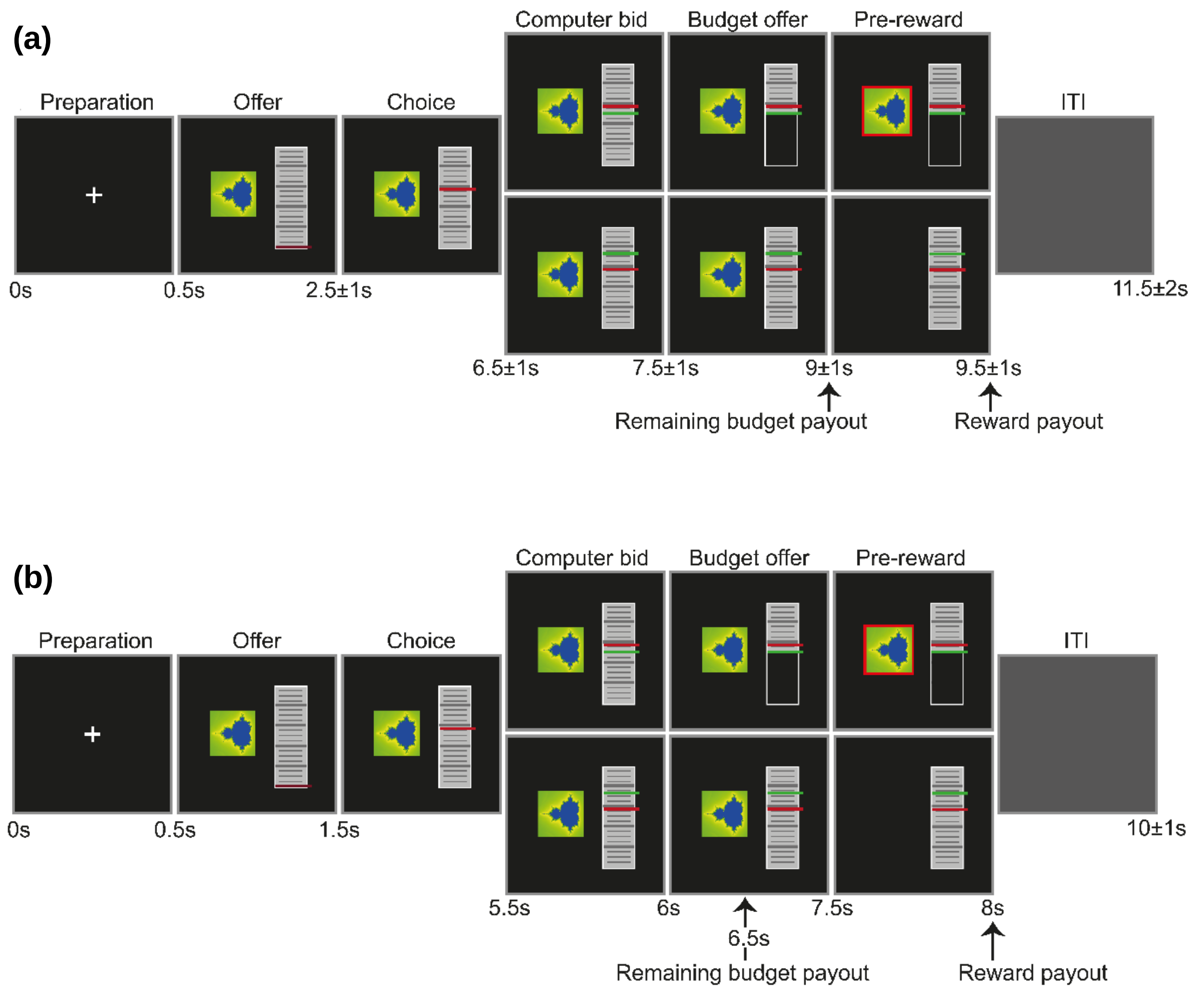


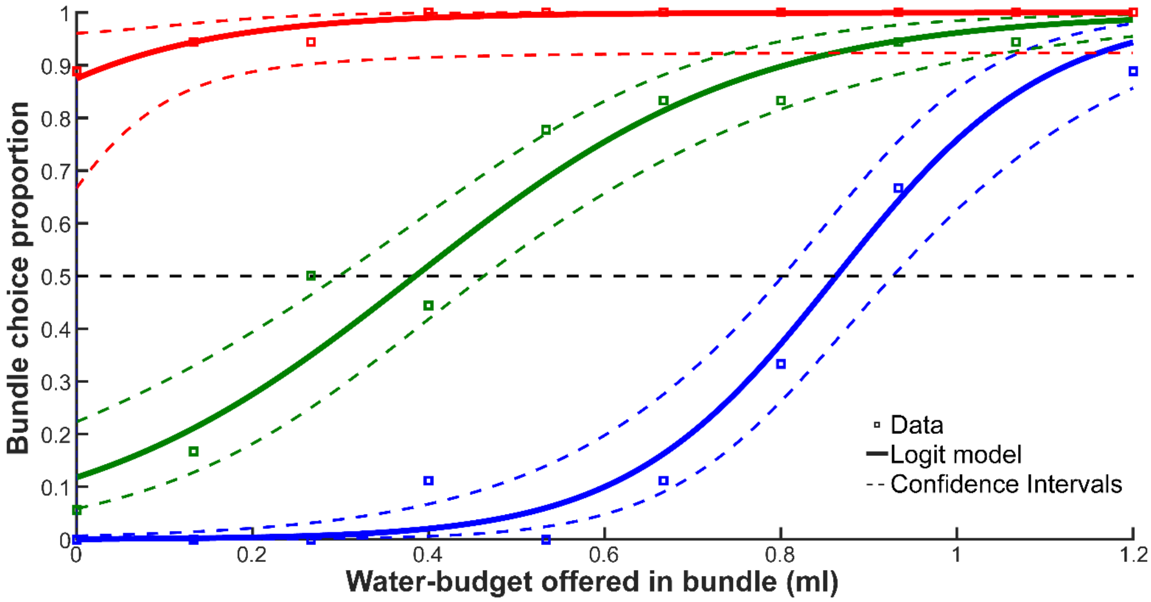
Fig. 5.26a) Trial structure of the [1.2-BDM-B] task, showing changes in the epoch timings, as well as the incorporation of a budget-bar with scale lines (as in the [1.2-BDM-A] version of the task). A 'jitter' was introduced to hamper the monkeys' ability to predict the timing of the 'choice' epoch without continuing to attend to the task. The length of the 'Computer bid' epoch was increased from 0.5s to 1s, increasing the presentation time of the outcomes to assist the monkeys in learning the task contingencies. Also, the delay between receipt of the remaining water-budget and receipt of the juice-reward was reduced from 1.5s to 0.5s – to increase the motivational value of the juice-rewards used in the task. The effect of the 'jitter' timing of ±1s is applied for all the remaining epochs, as before, the ITI duration is 2±1s, which leads to an overall trial end time of 11.5±2s when accounting for the effects of the 'jitter' alongside this variable ITI delay. **b)** The trial structure of the [1.2-BDM-A] task is shown for comparison. It used the same epoch timings as those that were applied when the binary-choice task was first introduced (Ch. 5.1) – these timings are identical to those of the preliminary BDM task (fig. 4.10), except for an increase in the choice time from 2s to 4s, and, are used for all BDM tasks that precede the [1.2-BDM-B] task in this chapter.

The reduced IRI of 0.5s was introduced to increase the value of the juice-reward relative to the water-budget, as a given volume of juice-reward should then be temporally discounted by a smaller amount. This would allow us to use lower volumes of the juice-reward, alongside the now larger water-budget volume of 1.2ml, and this could help to limit the rate at which the monkeys became sated. Thus, we now used smaller mango juice-rewards with volumes of 0.75ml, 0.6ml and 0.15ml – and intended to use 0.3ml and 0.45ml rewards in addition in a later version of the task if we saw suitably differentiated bids in this [1.2-BDM-B] task using just three rewards. These reward volumes were found to make the best use of the water-budget range across both monkeys (fig. 5.27), though the value of the 0.75ml reward for Ulysses was higher than the maximum water-budget, such that the optimal bid would be at the maximum, rather than the monkey's true value for the reward.

Initially, we had avoided such short IRIs as we wanted to limit any mixing of the juice-reward and water-budget in the monkeys' mouths, which could then be perceived as a different reward altogether, rather than a bundle of two separate rewards. Moreover, we speculated that such a separation in timing would emphasise the different stimulus-reward relationships in the task. Nevertheless, we now attempted to use a shorter delay as the monkeys had sufficient experience with the stimuli, and, as another ongoing task in the laboratory had shown that such a short IRI could be used when delivering bundles of rewards[79].

These timing changes led to dramatic improvements in the performances of both monkeys, in both the bottom-start, [1.2-BDM-B] (fig. 5.28) and top-start, [1.2-TS-BDM-B] (fig. 5.29), versions of the task. Both monkeys were trained in relevant target-tasks (Ch. 4.1) that were adapted to have equivalent reward-delivery timings to this new version of the BDM. Thus, after achieving >90% performance in a target-task with a bottom bid-marker starting position, they were trained in 10 sessions of the [1.2-BDM-B] task. Similarly, they had to achieve >90% performance in a top-start target-task before moving onto the [1.2-TS-BDM-B].

(a)



(b)

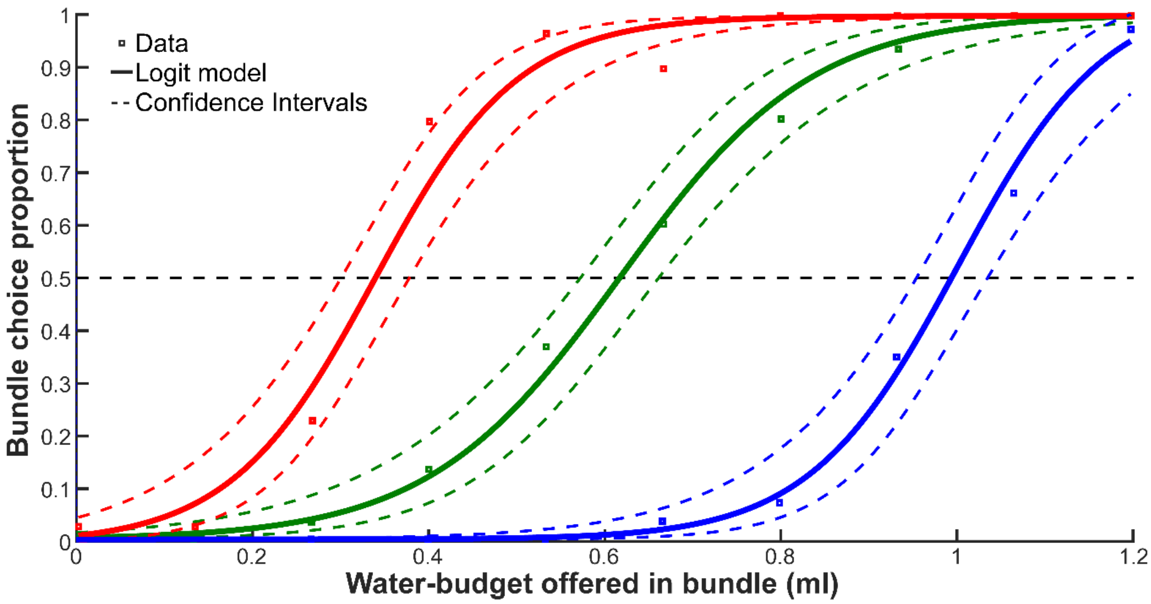
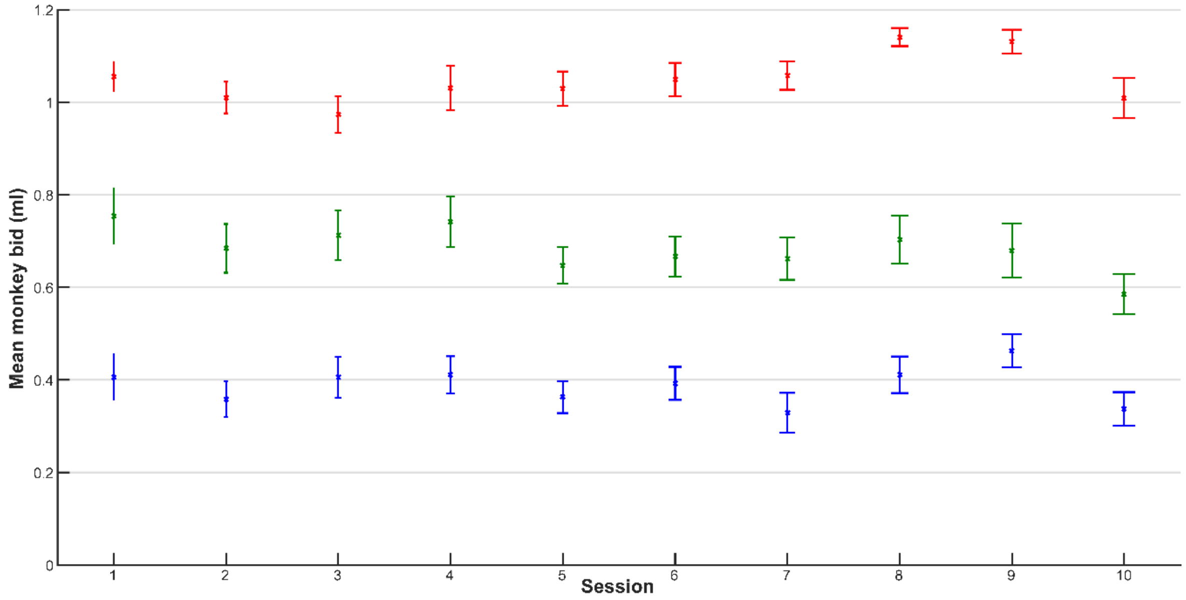


Fig 5.27a) The results of pooling choices over 3 sessions of the binary-choice-bundle (BCb) task in Ulysses, with equivalent reward-delivery timings to those shown in figure 5.26a for the BDM. The inferred optimal bids for the 0.75ml (red), 0.6ml (green), and 0.15ml (blue) mango juice-rewards were 1.2ml, 0.82ml, and 0.34ml respectively. Note that in the case of the BCb task, the equivalent bid is found by subtracting the amount of water-budget offered at the point of indifference from the total water-budget volume. **b)** The results of pooling choices over 6 sessions of the same task as shown in (a), but for Vicer. The inferred optimal bids for the 0.75ml (red), 0.6ml (green), and 0.15ml (blue) mango juice-rewards were 0.86ml, 0.58ml, and 0.20ml respectively.

(a)



(b)

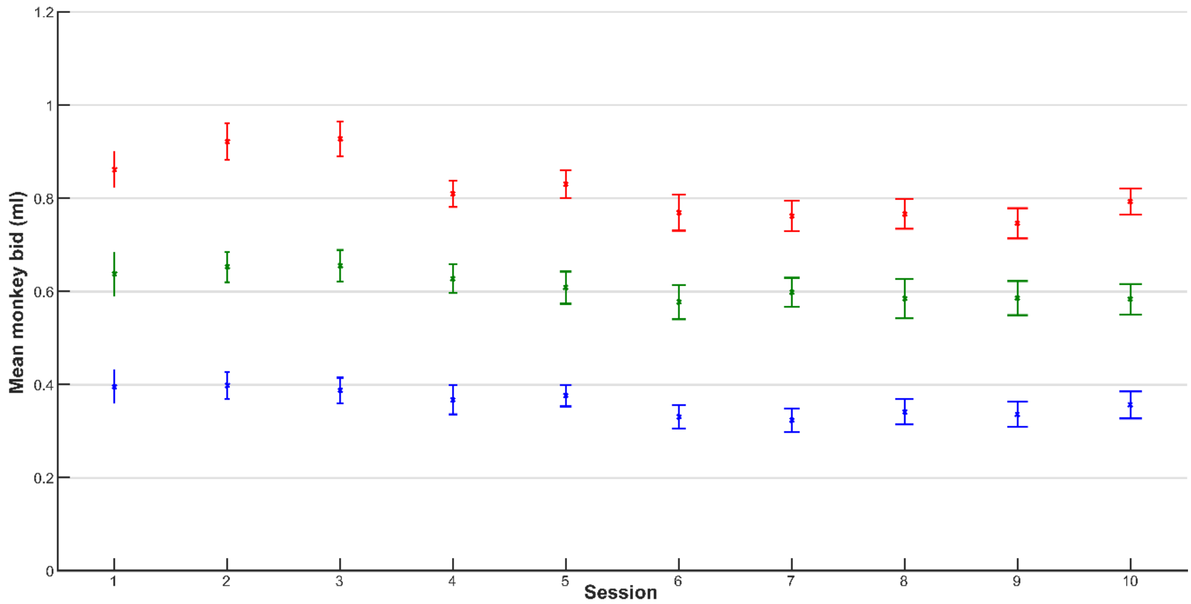
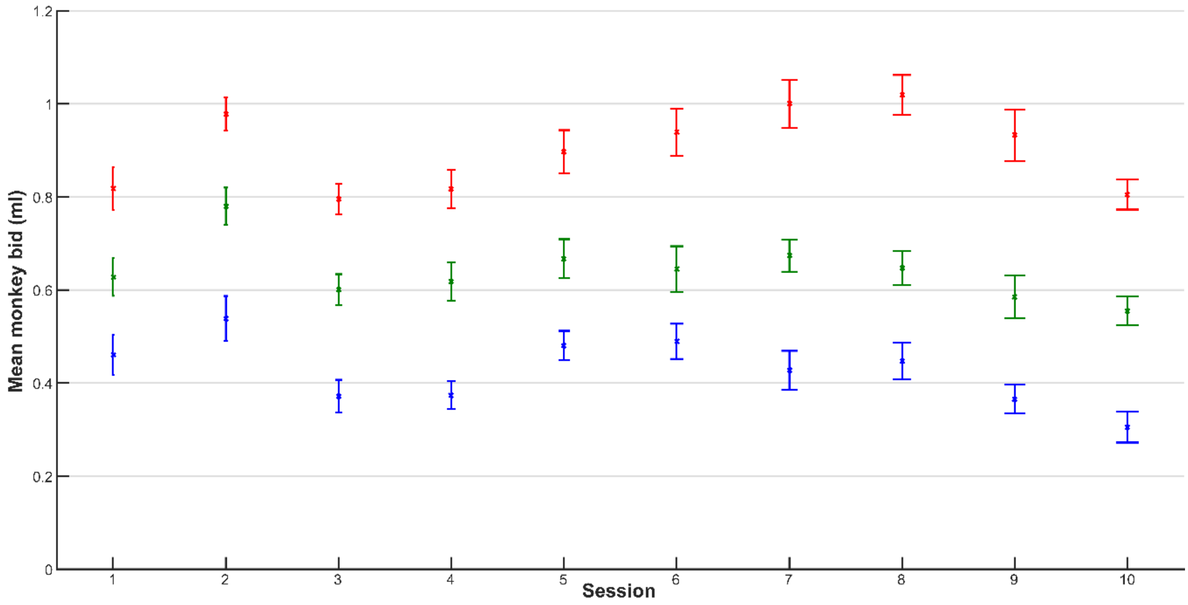


Fig. 5.28a) Mean bids of Ulysses in the [1.2-BDM-B] task for the 0.75ml (red), 0.6ml (green) and 0.15ml (blue) mango juice-rewards. Error-bars are 95% confidence intervals of the mean. **b)** As in (a) but for the second monkey, Vicer.

(a)



(b)

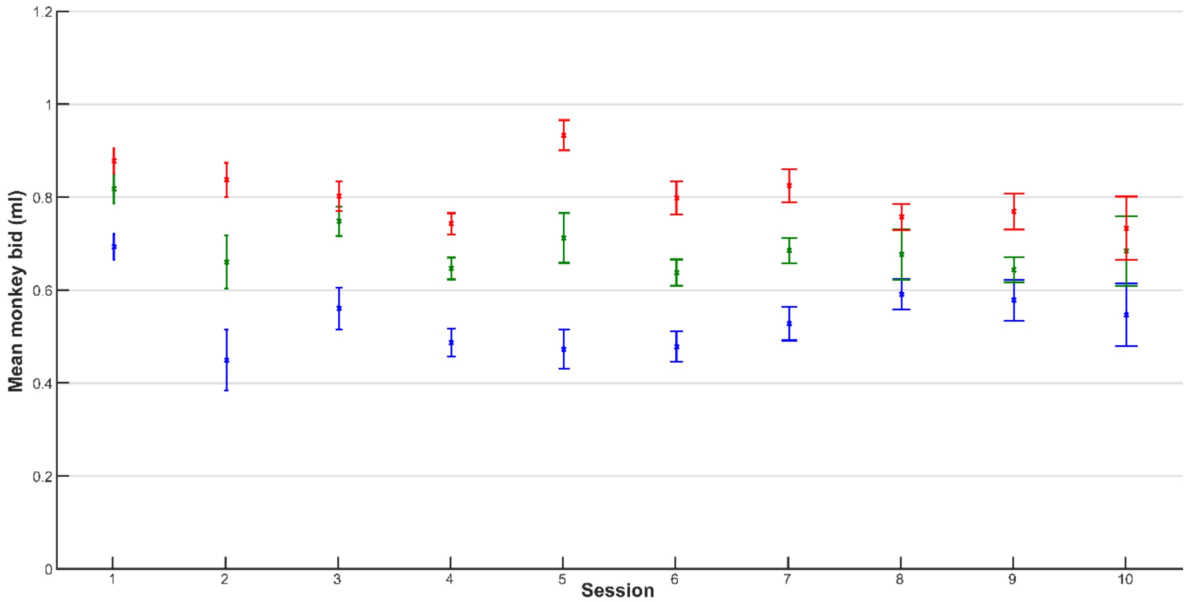


Fig. 5.29a) Mean bids of Ulysses in the [1.2-TS-BDM-B] task for the 0.75ml (red), 0.6ml (green) and 0.15ml (blue) mango juice-rewards. Error-bars are 95% confidence intervals of the mean. **b)** As in (a) but for the second monkey, Vicer.

1-way ANOVAs also found a significant effect of reward volume on the monkey's bids in every session of both the [1.2-BDM-B] and [1.2-TS-BDM-B] tasks in both monkeys (all $p<0.01$). For Ulysses, Bonferroni-corrected multiple-comparisons tests found a significant difference in the mean bids of all three rewards in each session (all $p<0.01$). This was also true of Vicer's bids in the bottom-start version of the task, and for most session in the top-start task (except for sessions 3 and 10). Thus, the monkeys' bids were consistently well differentiated and reflective of their preferences over the different volumes of mango juice-reward.

In the bottom-start, [1.2-BDM-B], task the mean Spearman's Rho for the correlation between the monkey's bids and reward magnitudes was 0.85 (SD = 0.031) for Ulysses and 0.84 (SD = 0.030) for Vicer. Both monkeys performed worse in the [1.2-TS-BDM-B] task: for Ulysses, the mean Rho was 0.80 (SD = 0.043) and for Vicer, mean Rho = 0.68 (0.15). This was to be expected given the greater wealth of experience that both monkeys had in various bottom-start versions of the BDM task. And, as Ulysses had some previous experience using a top-start BDM task (Ch. 5.1), it was unsurprising that his performance in the [1.2-TS-BDM-B] task would be superior to that of Vicer.

We can compare the accuracy of the monkeys' bids in different versions of the task by looking at the mean deviation from the optimal bid under each condition. By normalising this to the water-budget, we get a normalised absolute bid deviance (nABD)[*] from optimality, expressing the absolute distance of a bid from the optimum as a proportion of the total budget-range. This allows for comparison of bidding accuracy across tasks with different budget parameters.

For example, for Ulysses, when pooling across different rewards, the absolute bid deviance (ABD) in the 0.6ml water-budget SCD task (M = 0.16ml, SD = 0.084ml) is not statistically distinguishable from the ABD in the [1.2-BDM-B] task (M = 0.16ml, SD = 0.13ml). However, there is a significant difference in the nABDs of the two tasks (fig. 5.30a), with a greater proportional distance from optimality in the SCD task (M = 0.27, SD = 0.14) than in the [1.2-BDM-B] task (M = 0.13, SD = 0.11, $p = 3.11 \times 10^{-223}$). This was reflected in the costs that Ulysses incurred in each task

[*] Here, 'absolute' refers to the fact that this measure ignores whether bids were higher/lower than optimal bid, and only measures the magnitude of the difference between a bid and the optimum.

relative to an optimal bidder. A 1-way ANOVA comparing the effects of BDM condition (in SCD, [1.2-BDM-B] and [1.2-TS-BDM-B] tasks) on the costs that the monkey incurs relative to an optimal bidder finds a significant effect of the BDM task-type [$F(2,5740) = 36.74$, $p = 1.40\times10^{-16}$], and a Bonferroni-corrected multiple-comparisons test shows significant differences in the ECMs of all groups (all $p < 0.01$). Thus, Ulysses incurred significantly lower absolute costs in the [1.2-BDM-B] task than in the top-start version, [1.2-TS-BDM-B], or in the lower budget-volume SCD task (fig. 5.30b).

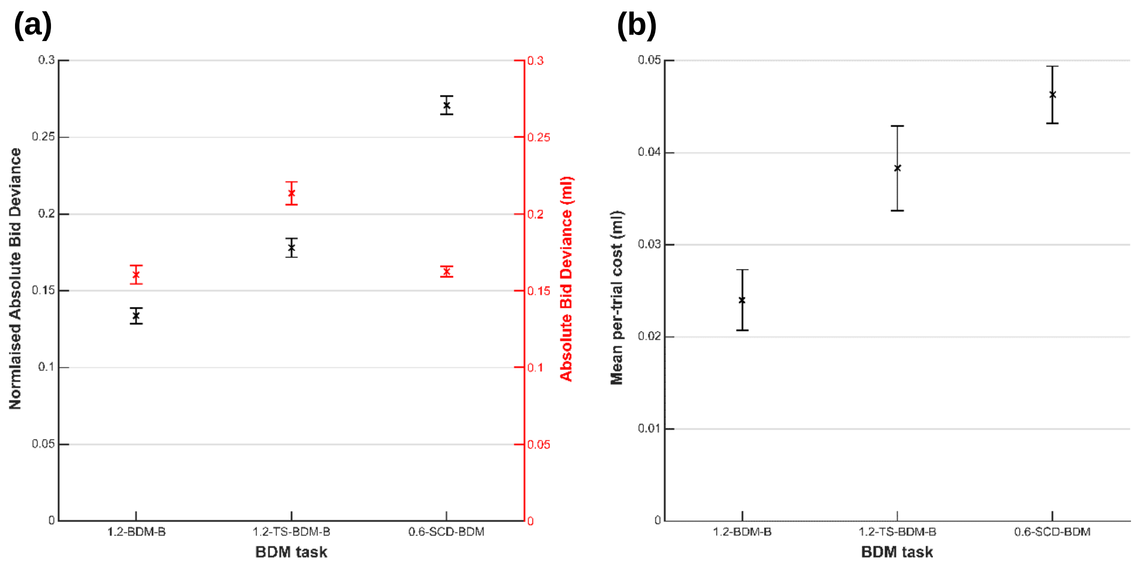
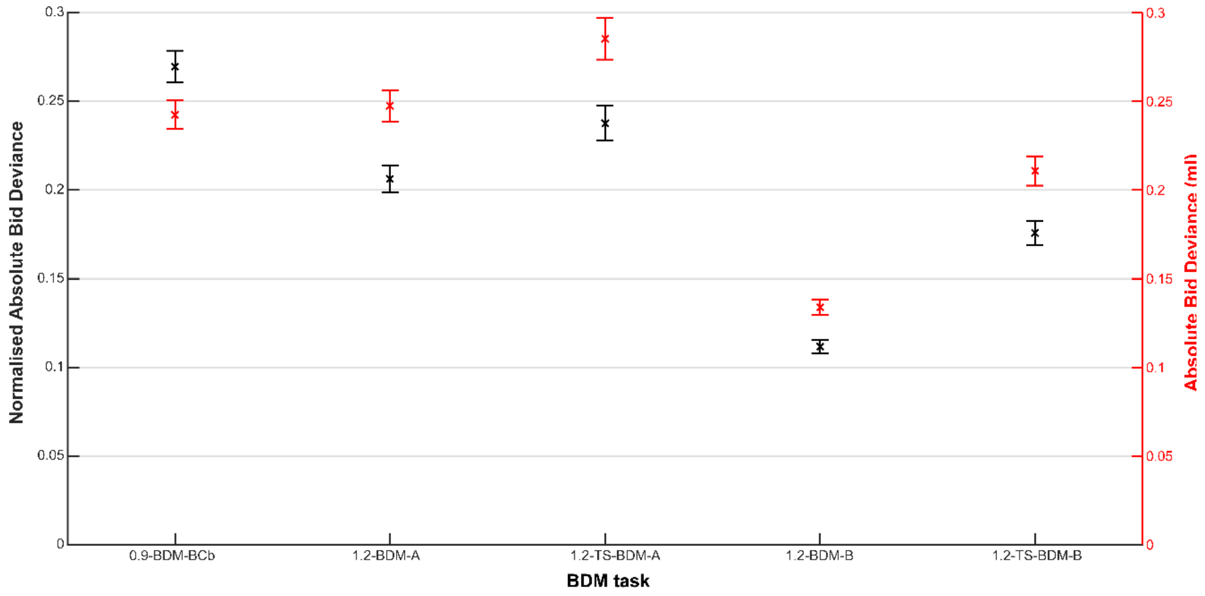


Fig. 5.30a) For Ulysses, the absolute bid deviance before (red) and after (black) normalisation is shown for each of three BDM tasks that use different parameters. On average, bids are proportionally closer to the optimum in the 1.2ml water-budget versions of the task, and is closest to optimality in the bottom-start condition, [1.2-BDM-B]. Error-bars are 95% confidence intervals of the mean. **b)** The mean per-trial cost is shown for each task in (a). This is calculated by finding the mean of the trial-by-trial differences between the monkey's payoff and the payoff that would be expected by a hypothetical optimal bidder facing the same computer bids. The mean per trial cost (ml) is expressed in terms of the water-budget numeraire. Error-bars are 95% confidence intervals of the mean.

Similarly, for Vicer, nABDs reflected a general improvement in performance as his training progressed through the various versions of the BDM described in this section (fig. 5.31a) – with his bids being relatively closer to the optimum in later versions of the task. Reflecting this, a 1-way ANOVA found a significant effect of BDM version on the mean per-trial costs [$F(4,8491) = 61.31$, $p = 3.79 \times 10^{-51}$] when comparing the effects of the [0.9-BDM-BCb], [1.2-BDM-A], [1.2-TS-BDM-A], [1.2-BDM-B], and, [1.2-TS-BDM-B] tasks (fig. 5.31b). Bonferroni-corrected multiple-comparisons tests showed significant differences in the mean per-trial costs between all task-types (all $p < 0.05$), except in the case of the [0.9-BDM-BCb] and [1.2-TS-BDM-B] tasks.

It is worth noting here, that despite the nABD of the [0.9-BDM-BCb] task being significantly higher than that observed in the [1.2-BDM-A] task, Vicer still incurred greater costs in the latter version of the BDM. This was possible because the reward values were relatively greater in the [1.2-BDM-A] (they were worth more proportional to the budget-volume being offered), and, therefore, the costs of a given deviation from the optimal bid were greater – indeed, this was our aim when introducing the higher water-budget volume, as these costs could drive learning over time.

(a)



(b)

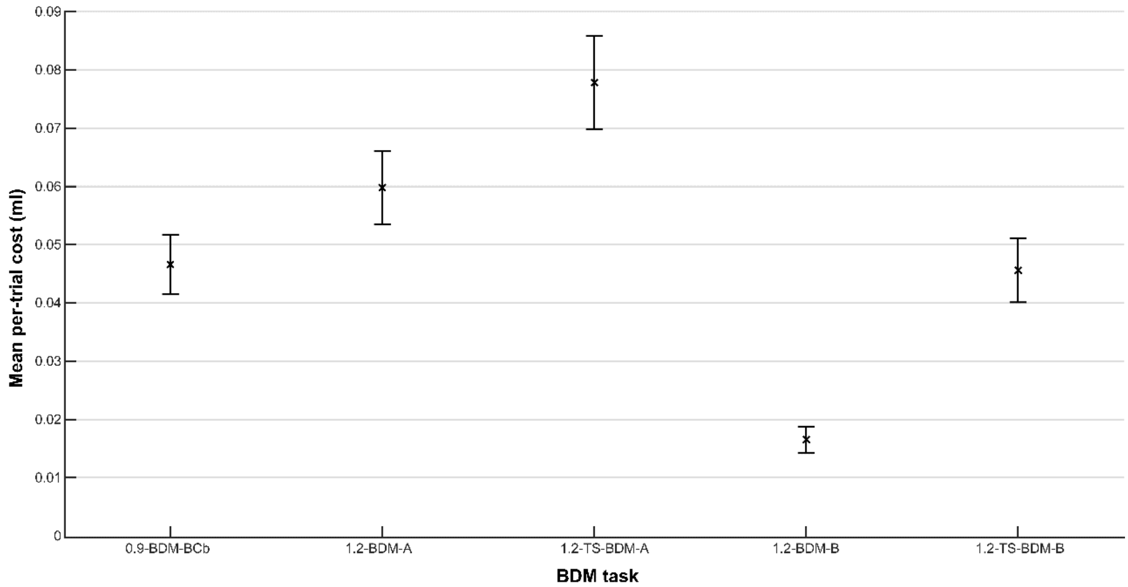


Fig. 5.31a) The absolute bid deviance before (red) and after (black) normalisation is shown for each of five key BDM tasks presented in this section, each testing different task parameters in Vicer. Bids are proportionally closer to the optimum in the 1.2ml water-budget versions of the task, and closest to optimality in the bottom-start condition, [1.2-BDM-B], as for the first monkey Ulysses. Error-bars are 95% confidence intervals of the mean. **b)** The mean per-trial cost is shown for each task in (a). This is calculated by finding the mean of the trial-by-trial differences between the monkey's payoff and the payoff that would be expected by a hypothetical optimal bidder facing the same computer bids. The mean per trial cost (ml) is expressed in terms of the water-budget *numeraire*. Error-bars are 95% confidence intervals of the mean.

Ulysses performed well using a random bid-marker start position:

The top-start (TS) position BDM task was introduced to rule out the possibility that the monkeys' responses were simply a product of increased response vigour on trials with larger magnitude juice-reward offers (Ch. 5.5). The performance of both monkeys in the [1.2-TS-BDM-B] task suggested that this was not the case. However, it was still possible that the monkeys had simply come to learn a reversed contingency between the reward magnitude and the effort to be expended in moving the joystick on a given trial, rather than intentionally making bids to specific budget-bar locations – that is, they may have learnt to move the bid-marker some specific distance, rather than understanding the significance of different points on the budget-bar. In that case, decoupling the distance between the optimal bid and the bid-marker start position would provide a more convincing demonstration of the fact that bids were being driven primarily by the reward value and an understanding of the task contingencies.

We had completed training of these tasks earlier in the first monkey, Ulysses, and so had time to test his performance in a random-start (RS) position version of the BDM task [1.2-RS-BDM-B]. First, Ulysses was trained in a random starting position target-task – i.e. a version of the target-task in which the bid-marker could appear in any position on the budget-bar at the start of the trial*. After achieving >90% performance in this target-task, his performance was tested in the [1.2-RS-BDM-B] task (fig. 5.32).

Overall, the correlation between Ulysses' bids and the reward magnitudes was comparable to that of the bottom-start version of this task: mean Rho = 0.85 (SD = 0.033). This was also reflected in the fact that the mean bids for all rewards groups were significantly different to one another in each session, as shown by a Bonferroni-corrected multiple-comparisons test (all $p < 0.01$). Moreover, of the three tasks in this section for which Ulysses was tested with a 1.2ml budget, he incurred the smallest mean per-trial cost in this random-start version of the task (M = 0.038, SD = 0.10) – though this was not significantly less than the mean per-trial cost incurred in the [1.2-BDM-B].

* See Chapter 4, Section 1, for a description of the target-task. This was simply adapted by matching the stimuli and reward volumes to those of the [1.2-BDM-B] task and randomly generating the bid-marker's start position on the budget-bar on every trial.

These results provided the most convincing example of rational bidding behaviour thus far. The monkey's bids remained consistent despite their inability to apply a specific learnt action in response to a given juice-reward predicting stimulus in this version of the task. We would go on to test the behaviour of both monkeys in bottom-start (BS), top-start (TS) and random-start (RS) versions of the final BDM task.

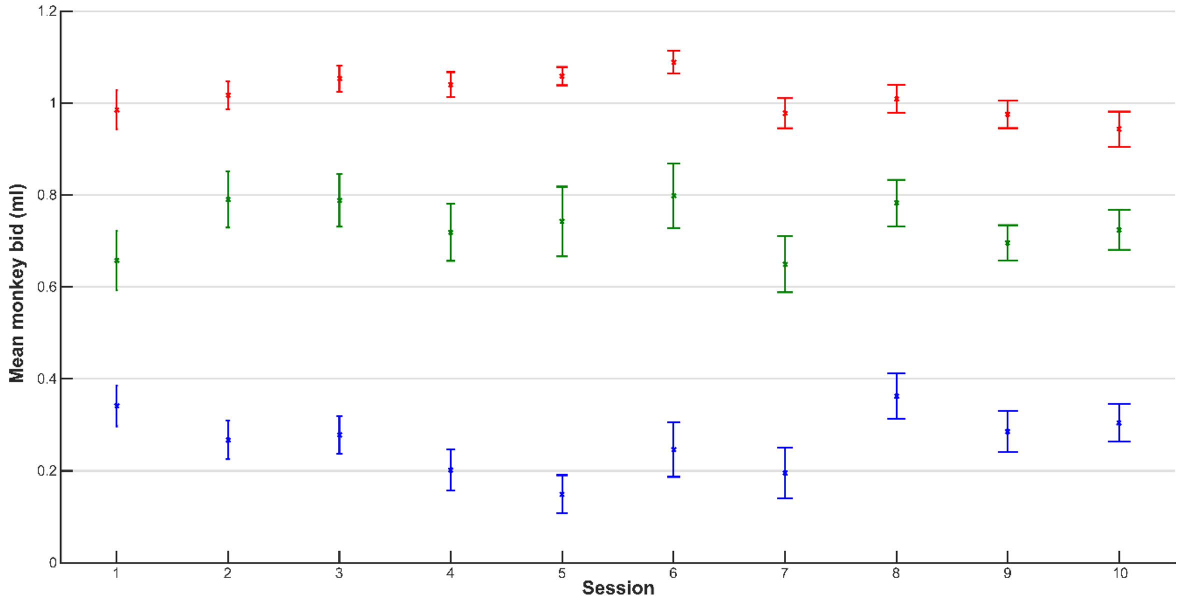


Fig. 5.32) The mean bids of Ulysses in 10 sessions of the random start-position version of the [1.2-BDM-B] task. Mean bids are shown for the 0.75ml (red), 0.60ml (green), and 0.15ml (blue) mango juice-rewards. Error-bars are 95% confidence intervals of the mean.

Summary of task performance:

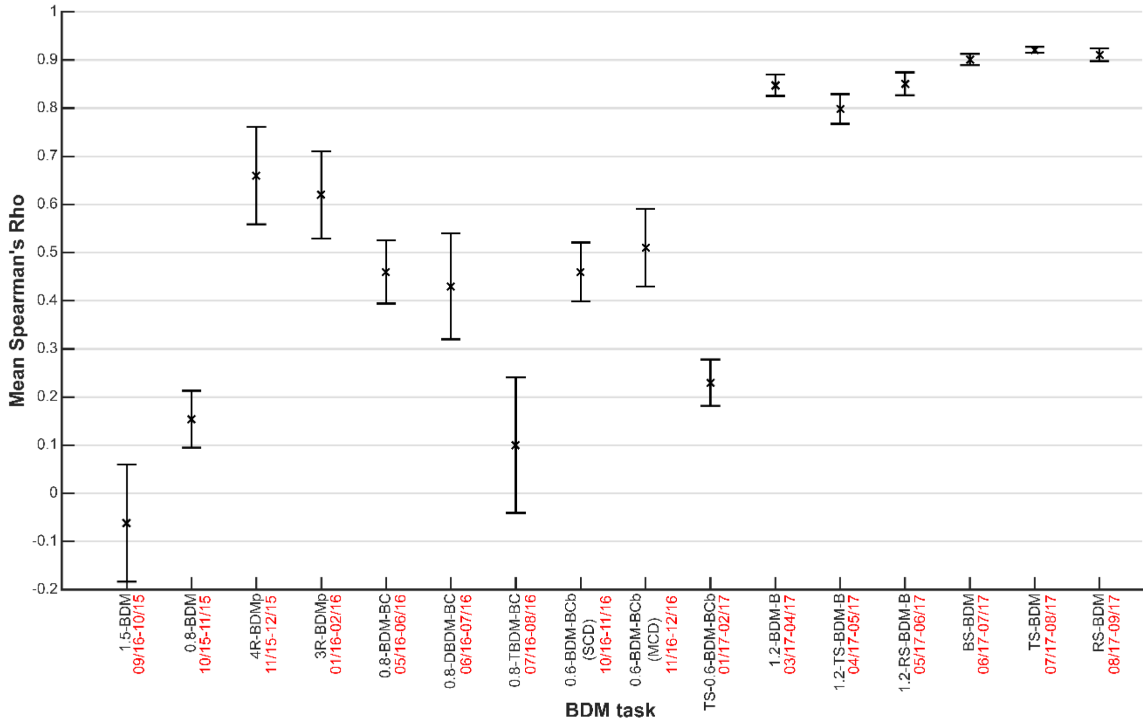
A dramatic improvement in the performances of both monkeys was observed after increasing the budget volume, training the BDM and BCb tasks on separate days, and, applying certain task-timing changes to increase the monkeys' motivation and attention.

We would apply two further timing changes to the final version of the BDM task, which would use 5, rather than 3, volumes of juice-reward (fig. 2.7): an increase in the choice time from 4 seconds to 6 seconds, and an increase in the inter-trial-interval (ITI) from 2±1s to 4±2s. The choice time was increased to reduce the urgency of decision-making (particularly important given the additional demands of using 5 juice-rewards), and, to improve the accuracy of bid-marker movement. While the increased ITI should enhance the rate of learning by making each trial more impactful[80,81]. Apart from these timing changes and the addition of two new reward volumes, the final version of the BDM task, presented in Chapter 2, was identical to the [1.2-BDM-B] task that was outlined earlier in this section.

Here, we present the results of all the BDM tasks described in this thesis for both monkeys. As in previous analyses, we take the Spearman's Rho for the correlation between the monkeys' bids and the reward magnitudes as our measure of performance, as it captures the consistency with which the ordering of the monkeys' bids reflects their order of preference over the different reward magnitudes that were tested.

For each monkey, we show the mean value of Spearman's Rho for all sessions of that version of the task (fig. 5.33). The different task-types are presented in the order in which they were tested, starting with the results of our preliminary version of the BDM (Chapter 4), then presenting the various intermediate versions of the BDM introduced in this chapter, and, concluding with the results of the final BDM task (Chapter 3) for comparison. Tables are also provided to match task names to the sections of the thesis where they are first presented (Table 5.4 for Ulysses and Table 5.5 for Vicer).

(a)



(b)

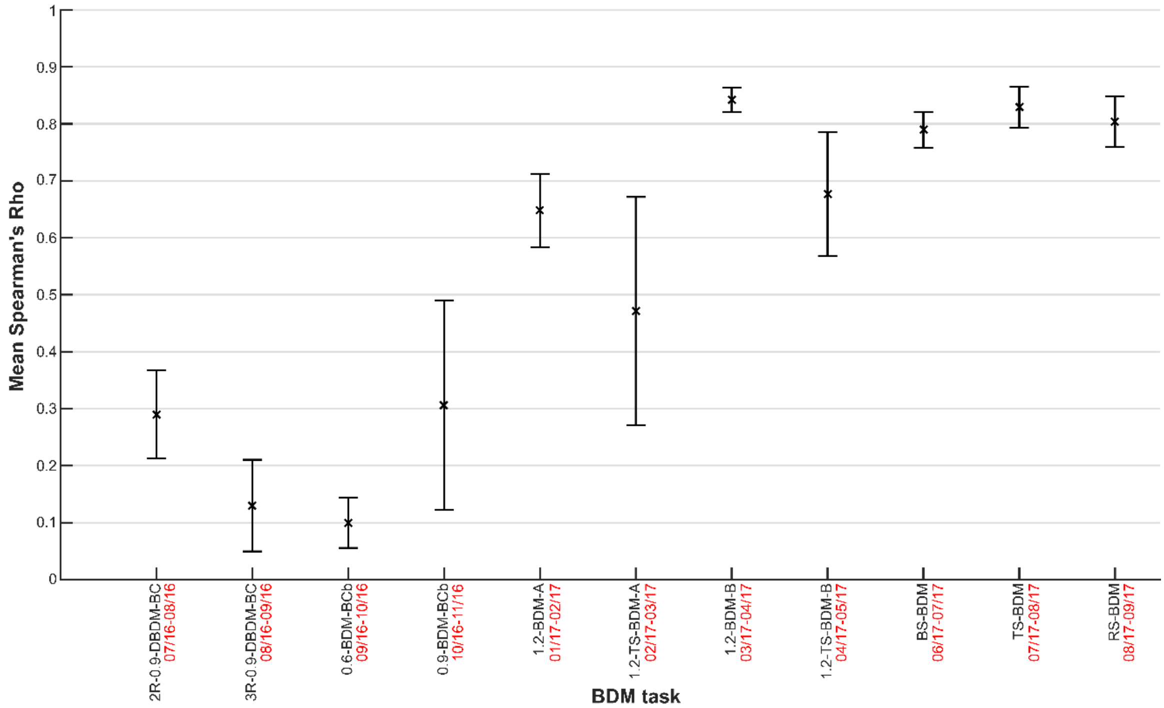


Fig. 5.33a) The mean Spearman's Rho for the session-by-session correlation of reward magnitude and the monkey's bid for each BDM task tested in Ulysses. Error-bars are 95% confidence intervals of the mean. Tables 5.4 and 5.5 refer to where each task is first described. The month within which a condition was tested is shown in red to the right of its name. **b)** As for (a), but for the second monkey, Vicer.

Table 5.4) Summary of each task and a reference to where each task tested in Ulysses can be found in the thesis. Also given are the number of sessions (NS), the mean Spearman's Rho across all sessions (MSR) and the standard deviation (SD) of Spearman's Rho across all sessions.

Task	Chapter	NS	MSR	SD	Figure
1.5-BDM	4.2	10	-0.062	0.17	4.11
0.8-BDM	4.2	13	0.15	0.098	4.13
4R-BDMp	4.2	6	0.66	0.096	4.15
3R-BDMp	4.2	6	0.62	0.086	4.16
0.8-BDM-BC	5.1	13	0.46	0.11	5.3
0.8-DBDM-BC	5.2	9	0.43	0.14	5.7
0.8-TBDM-BC	5.2	6	0.1	0.13	5.10
0.6-BDM-BCb (SCD)	5.4	10	0.46	0.085	5.16a
0.6-BDM-BCb (MCD)	5.4	10	0.51	0.11	5.16b
0.6-TS-BDM-BCb	5.5	15	0.23	0.087	5.22
1.2-BDM-B	5.6	10	0.85	0.031	5.28a
1.2-TS-BDM-B	5.6	10	0.80	0.043	5.29a
1.2-RS-BDM-B	5.6	10	0.85	0.033	5.32
BS-BDM	3.1	10	0.9	0.016	3.1
TS-BDM	3.1	10	0.92	0.0078	3.1
RS-BDM	3.1	10	0.91	0.018	3.1

Table 5.5) Summary of each task and a reference to where each task tested in Vicer can be found in the thesis. Also given are the number of sessions (NS), the mean Spearman's Rho across all sessions (MSR) and the standard deviation (SD) of Spearman's Rho across all sessions.

Task	Chapter	NS	MSR	SD	Figure
2R-0.9-DBDM-BC	5.2	5	0.29	0.062	5.8
3R-0.9-DBM-BC	5.2	5	0.13	0.065	5.9
0.6-BDM-BCb	5.6	10	0.1	0.062	N/A
0.9-BDM-BCb	5.6	10	0.31	0.26	5.23
1.2-BDM-A	5.6	10	0.65	0.090	5.24b
1.2-TS-BDM-A	5.6	10	0.47	0.28	5.25
1.2-BDM-B	5.6	10	0.84	0.030	5.28b
1.2-TS-BDM-B	5.6	10	0.68	0.15	5.29b
BS-BDM	3.1	10	0.79	0.044	3.3
TS-BDM	3.1	10	0.83	0.050	3.3
RS-BDM	3.1	10	0.80	0.062	3.3

6

Making use of the BDM in monkeys

The development of the Becker-DeGroot-Marschak (BDM) task for monkeys requires a relatively long and complex multi-stage training period (Ch. 4.1). The monkeys' bids successfully reveal their preferences over 4-5 rewards within a 200-trial session (Ch. 3.1), but deviate from the values of those rewards as inferred in an analogous binary choice (BC) task (Ch. 3.2). Nevertheless, the disadvantages of this method are compensated for by the wealth of novel information present in the monkey's responses, and the wealth of information that they must integrate to produce those responses.

Apart from providing a measure of subjective value at a greater temporal resolution than that achieved in standard BC tasks, the BDM also imposes a clear conceptualisation and quantification of relevant costs, as well as a dissociation between choices and outcomes - contrast this with the unambiguous outcomes of BC tasks, even those which utilise risky rewards, or lotteries.

The BDM continues to be used as a staple method of value elicitation in behavioural economics, and is growing in popularity in neuroscientific studies of human decision making; being far more practicable and efficient than BC tasks. Of course, experiments that make use of the method are far more easily developed for and executed by human subjects, but a monkey analogue will only become increasingly useful as cross-species comparisons of methods continue to bear fruit - see, for example, the adaptation of Hernadi et al.'s (2015) monkey saving task[82] by Zangemeister et al. (2016) for use in a human fMRI experiment[83].

Eliciting values for their own sake is sufficient reason to use the BDM for a market researcher, but key to its justification as a practical tool in the neuroscience laboratory are the potential future experiments that would make use of the 'richness' of the method and allow for more meaningful comparisons with similar studies in humans.

This chapter outlines a neurophysiological experiment in monkeys that would benefit from utilising the BDM.

1 – Neurophysiology in the BDM task

One of the most useful aspects of the BDM is its ability to elicit a trial-by-trial response from the subject that reflects their subjective value for the reward being offered. In current BC tasks, the subject's choices on any given trial should reflect their subjective values, but doesn't necessarily provide a behavioural response that is itself a direct expression of the chosen offer's value. The BDM provides a trial-by-trial value signal in the form of a bid, whilst in BC tasks both values and preferences must be inferred by analysing choices over dozens of trials that are temporally disjointed.

Current understandings of value coding in the orbitofrontal cortex (OFC) as well as in the midbrain dopamine (DA) neurons of the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc), have relied upon combining neuronal data with post-hoc inference of values[3,30]. The use of such simple methods is well suited to neurophysiology, both by reducing noise in the behavioural responses of the subject and by making data analysis simpler. However, the BDM provides an opportunity to correlate behavioural responses with neuronal activity in a way that that makes the most of the high temporal resolution of extracellular recordings.

This issue of temporal specificity speaks to some of the earliest concerns that drove the initial development of the BDM method. Describing earlier attempts to derive the utility of money using a BC task[32], Becker et al. (1964) drew attention to the fact that such a method "depends heavily on the assumption that the subject's probabilities of choice remain constant throughout the many times that he is choosing from the same available set of actions", and, earlier in the same paper[1], they critique the use of stochastic models to infer values on the basis that "it is assumed that these probabilities do not change during the time period under consideration, thus precluding learning or any systematic change of behavior". Indeed, if such temporal imprecision presents a problem for the behavioural economist then it is likely to be problematic on the millisecond scale of extracellular recordings. For example, the BC task requires the pooling of choices over dozens of trials to allow the experimenter to infer a single subjective value, making less use of the evolving value signals that are elicited at the level of the neuron by using a behavioural task with low temporal resolution. On the other hand, the BDM elicits a value on each trial and more closely matches the temporal resolution of neurons themselves, potentially allowing for the identification of neuronal correlates of behaviour as it changes from trial to trial.

The development of the task has required the training of subjects in both the BDM and BC tasks. It is therefore possible that neuronal recordings could be used to investigate whether various neuronal value signals are independent of the mechanisms used to elicit them. Monkeys trained in the BDM would be uniquely capable of valuing reward offers in two different incentive compatible tasks, potentially within the same recording session. With regards to this, it is important to note that bids in the BDM task are produced by an action that directly represents the subject's value for the reward offer, while choices in BC tasks dissociate value and action insofar as the end goal of the action is not itself an expression of the subject's value.

This distinction leads to certain neurophysiological recording disadvantages for the BDM relative to the BC task, whereby dissociation of action and subjective value in the latter could make it easier to disambiguate the correlates of a given neuronal signal. On the other hand, if there is a functionally significant difference between those actions that specifically report an 'analogue' value and those that are simply driven by subjective values, then a comparison between the two tasks could allow for its identification.

One could reasonably speculate that the striatum may be a good candidate region for the exploration of such relationships between action and value[84] by comparing neuronal activities during BDM and BC task performance. There are a multitude of such potential recording sites and experiments, and detailing all such interesting possibilities is beyond the scope of this thesis. Instead, this section focuses on three candidate regions for neurophysiological recording: Midbrain DA neurons, value-coding neurons of the OFC, and the prediction error (PE) coding neurons of the lateral habenula (LHb).

<u>Modifying the BDM for neuronal recording experiments:</u>

The strength of the BDM lies in the wealth of information that the subject is responding to throughout the task, and its ability to produce a trial-by-trial 'analogue' report of subjective value. However, this same strength is in another sense a potential source of weakness in the context of extracellular recordings. The wealth of information that must be presented on-screen, and updated throughout the task, could make it difficult to disentangle value and other decision related signals from those that simply reflect the changing stimuli when analysing noisy neuronal data.

The BDM task will therefore have to be adapted to minimise the effects of such confounds, reducing the complexity of the task such that responses to key stimuli can be more easily identified, and removing 'redundant' epochs to minimise the length of each trial - which, at 15.5s (Ch. 2.2), may prove problematic given the difficulty of isolating task-relevant neurons for extended periods of time.

When training monkeys in the BDM these additional stimuli and epochs may have encouraged learning by more clearly signifying changes during the trial, but it may be possible to remove them once the monkeys have shown evidence of robust performance of the task.

Such an adapted experimental design is outlined in figure 6.1. Firstly, the reward and budget stimuli are now presented sequentially, so that neuronal responses to each of these can be better isolated: the reward is shown first in a 'reward offer' epoch, and following a short delay the budget and marker starting position are shown in the 'bid preparation' epoch. The delay of 0.5s provides a time-window within which neuronal responses to the reward stimulus can be isolated (a 350ms time-window has been shown to be sufficient to identify responses to individual conditioned stimuli in amygdala neuron recordings, for example[85]). Responses that are dependent upon the subjective value of the reward should correlate with the monkeys' bids.

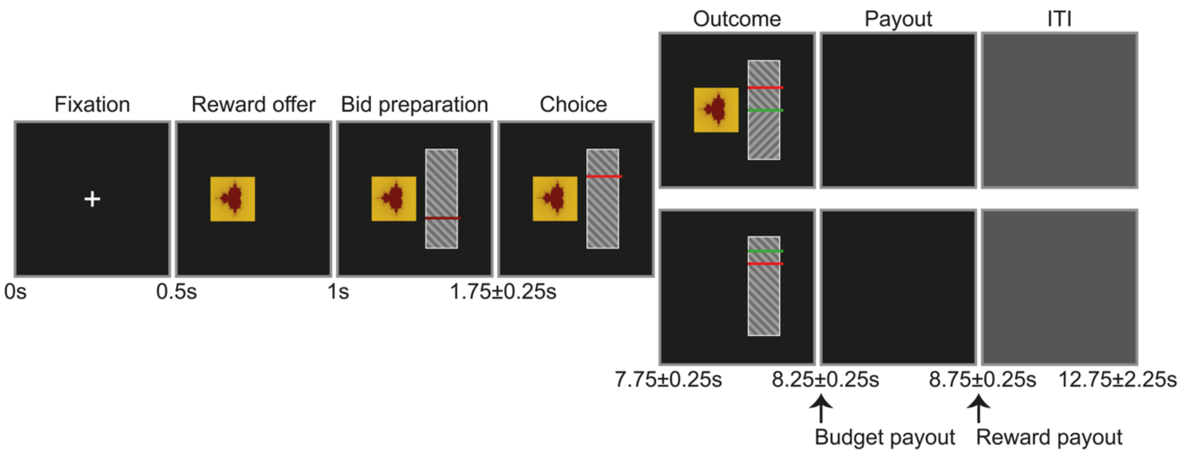


Fig. 6.1) An adapted version of the BDM task for use during neuronal recordings. A 0.5s exponentially distributed jitter is applied between the 'bid preparation' and 'choice' epochs, and therefore effects all following epoch timings. Another 2s jitter is applied during the ITI.

Furthermore, both the delay and jitter between the 'bid-preparation' and 'choice' epochs have been reduced so that the overall trial time is reduced but the start of the bidding epoch is still not perfectly predicted. This has two intended effects, first to maintain attention, and second, to provide a separation between movement preparatory activity and responses to the bid-preparation epoch. The 'bidding' epoch itself has remained unchanged with a total of 6s allocated for bidding. However, this could be another epoch whose length could be reduced to shorten total trial time if it is found that the monkeys' performance is not too adversely affected.

Finally, the 'outcome' epoch now combines the presentation of the computer-bid and the budget payment into a single stimulus change, as well as the disappearance of the reward fractal in cases where the subject has made a losing bid. These changes are intended to remove information redundancy, as the overall trial outcome can be determined by the relative locations of the monkey and computer bids. This allows for further reductions in trial time and the number of on-screen stimulus changes. Moreover, instead of shading the paid portion of the budget-bar, this region is now covered by a grating with an opposite direction to that of the rest of the bar. Thus, the amount of budget to be paid is signalled in a luminosity preserving manner, maintaining the same level of stimulus intensity, to which DA neurons of the VTA, for example, are known to be sensitive[86].

Ultimately, changes to the task design will have to be made alongside the recording of neurons during task performance; allowing the experimenter to make changes based on the monkeys' capabilities, and as yet unknown responses of neurons. Nevertheless, the outline of changes presented here can counter some of the tasks weaknesses and make it more suitable for neurophysiological recording experiments.

Candidate regions for neuronal recording in the BDM:

Midbrain Dopamine -

Of immediate interest for our laboratory would be a replication of previous findings of the representation of value information in the form of PEs coded by DA neurons of the SNc and VTA[87,88]. Within the BDM task there are several potential influences on the PE signal, and several different PEs that could be experienced and expressed over the course of a BDM trial, allowing us to observe the integration and relative contribution of these various influences on the PE signal over time.

The responses of DA are expected to be of particular interest at two key epochs within the task. Firstly, in the response to the first presentation of the unpredicted reward in the 'reward offer' epoch (fig. 6.1). The magnitude of the PE during this epoch should be related to the subjective value of, and therefore the mean bid for, that reward - allowing a relationship to be drawn between a monkey's behaviourally reported value and the value implied by DA coding of the PE.

The next important PE to observe is expected to occur during the 'outcome' epoch (fig. 6.1) - at this point in the task the overall reward outcomes are revealed for both the reward and budget liquids. Just before this, during the 'bidding' epoch, the subject is hypothesised to have an expectation of the overall value of the outcome based upon their knowledge of the computer's bidding distribution. If the subject understands the contingencies of the task, then their expected total value at the end of the 'bidding' epoch, expressed in terms of the water-budget as numeraire is as follows:

$$E[V]_{Bidding} = p(win)(budget - E[Cbid] + reward\ value) + p(lose)(budget)$$

Where both ***reward value*** (the value of the juice-reward offer) and ***E[Cbid]*** (the expected computer bid) are expressed in millilitres of water-budget.

Because the probability of winning in the BDM, ***p(win)*** is simply the probability that the subject's bid is greater than the computer's bid, then both ***p(win)*** and ***E[Cbid]*** are determined by the subject's knowledge of the computer-bid distribution (the probability of losing, ***p(lose)***, is simply equal to ***1 – p(win)****)*. Therefore, the PE that would be expected at the beginning of the 'outcome' epoch - when the subject has the remaining budget revealed to them and finds out whether they won the juice-reward– is simply determined as:

$$PE_{Outcome} = Value_{Outcome} - E[V]_{Bidding}$$

The DA PE signal is robust and it has already been shown that DA responses are modulated in a way that reflects integrated subjective values and multiple rewards, that can be expressed in terms of a numeraire, rather than in terms of individual reward attributes[27,70]. Therefore, as well as allowing for trial-by-trial testing of the developing PE signal in DA neurons, identification of this expected PE coding would provide a reliable test for assessing the viability of neurophysiological recording within the BDM task.

Orbitofrontal cortex –

Another candidate region for recording in this task is the OFC, particularly areas 11 and 13, with an emphasis on more anterior regions that are thought to be involved in the abstract coding of subjective value[89]. Activity in this region has been correlated with subjective value in both human fMRI experiments[90], some of which utilised the BDM to elicit this activity[14], as well as in neurophysiological recordings in monkeys[91].

Apart from replicating those key findings of value representation, or examining the mechanism dependence/invariance of these signals, several other issues may be simultaneously explored in the OFC while subjects perform the BDM.

For example, do neurons of the OFC represent the values of the separate components of the BDM (reward and budget liquids), or do they represent the overall subjective offer value of the trial? Do neurons of this region account for expected costs within the task, based on an understanding of the computer's bidding distribution – i.e. do they integrate the expected cost into the object value representation of the budget liquid, or represent the value of the reward in a way that reflects the probability of winning it? Or, are these values updated later, after the outcomes of the trial are revealed?

By taking advantage of the developing and changing predictions and values throughout a BDM trial, we can query the integration of different sources of task-relevant information and their role in guiding the subject's value-based response.

Lateral Habenula -

Activity of the glutamatergic neurons of the lateral habenula (LHb) has been correlated with both positive and negative prediction errors[92], with excitation in response to negative PEs and suppression of activity in response to positive PEs – the opposite pattern to that which is observed in DA neurons of the midbrain (fig. 6.2). Moreover, in trials where a negative PE is produced, the excitatory responses of neurons in the LHb precedes the inhibitory responses of DA neurons of the VTA, suggesting that the activity of these habenular neurons could be driving the suppression of DA activity that occurs following a negative PE.

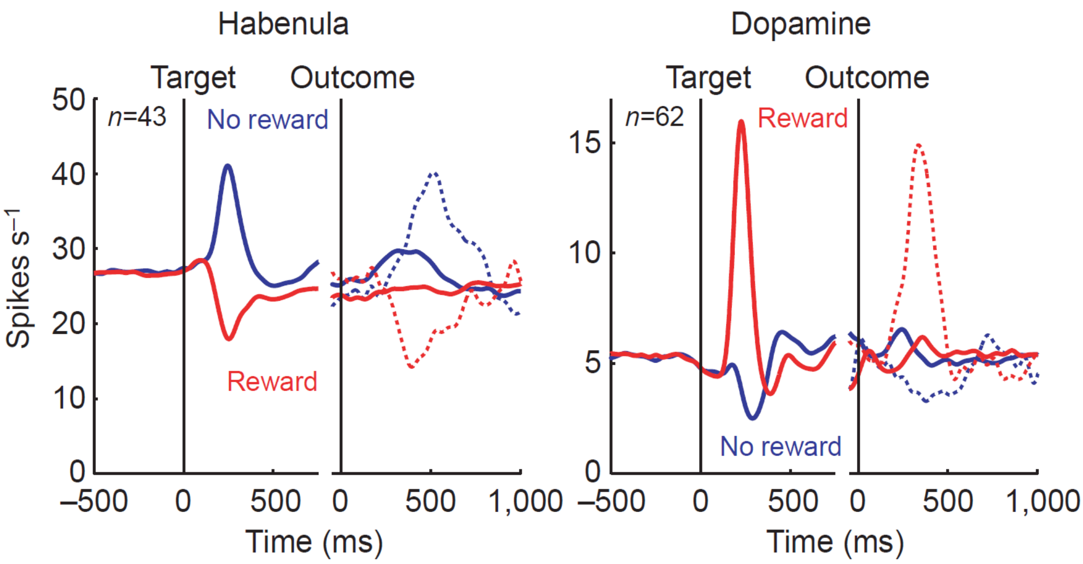


Fig. 6.2) Averaged activity of LHb and VTA DA neurons showing modulation in opposite directions by conditioned stimuli that either predict a reward or its omission (adapted from figure 3a of [92]).

The organisation of LHb-VTA/SNc projections is consistent with such a polar relationship between the two structures, with excitatory glutamatergic neurons of the LHb forming a disynpatic circuit with the VTA/SNc via inhibitory GABAergic neurons of the rostromedial tegmental nucleus (RMTg). Consistent with this architecture is the finding that GABAergic RMTg neurons are also activated in response to aversive stimuli and inhibited by those that signal reward[93].

Matsumoto and Hikosaka went on to show that LHb neurons respond to ‘negative motivational value’, having graded excitatory responses to conditioned stimuli that predict aversive outcomes with greater certainty[94], and find the opposite pattern in LHb neurons in response to appetitive outcomes (fig. 6.3). They later showed a link between LHb stimulation and behaviour by observing the effects of electrical stimulation of the LHb while NHPs learnt a visually guided saccade task – whereby their findings suggested that “LHb activity contributes to learning to suppress actions which lead to unpleasant events”[95].

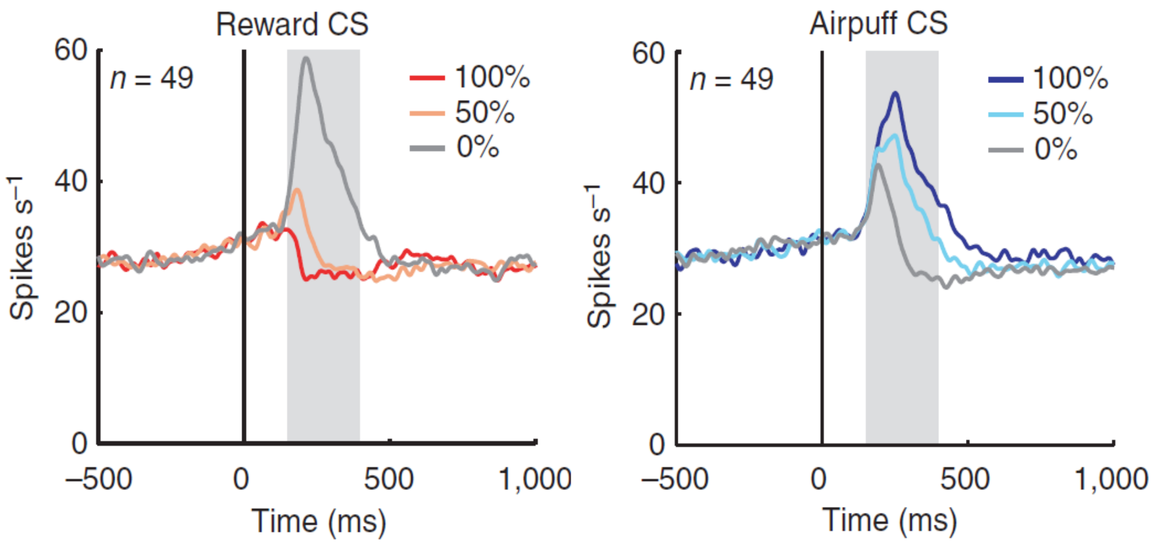


Fig. 6.3) Averaged activity of neurons in the LHb in response to rewarding and aversive CSs signalling the outcome at different probabilities, with spike density functions for each shown in different colours (adapted from figure 2c and 2d of [94]).

These findings point to the LHb playing a behaviourally relevant role in learning, and alongside the expression of PEs and its anatomical connections with the DA reward system, there is reason to believe that the LHb plays an important role in the assessment of negative outcomes for decision making. Nevertheless, despite these findings, the various experiments conducted still only provide a qualitative account of the LHb's coding of reward-related information - there has still been no quantitative assessment of the coding of negative value signals in this region, nor has there been any attempt to relate this to the midbrain DA signal within an economic framework.

It is in this regard that recording neurons of the LHb during performance of the BDM could provide new insights. Concurrent recordings from the LHb and VTA DA neurons would allow for the observation of opposing PE signals at different phases of the task, with quantification of the values that the subject experiences also being provided on each trial. More specifically, the BDM task allows for manipulation of expected costs by changing the computer-bid distribution – this can be signalled to the subject such that the magnitude of the PE response can be correlated with a quantifiable difference between the value that the subject experienced and the one that they expected (fig. 6.4).

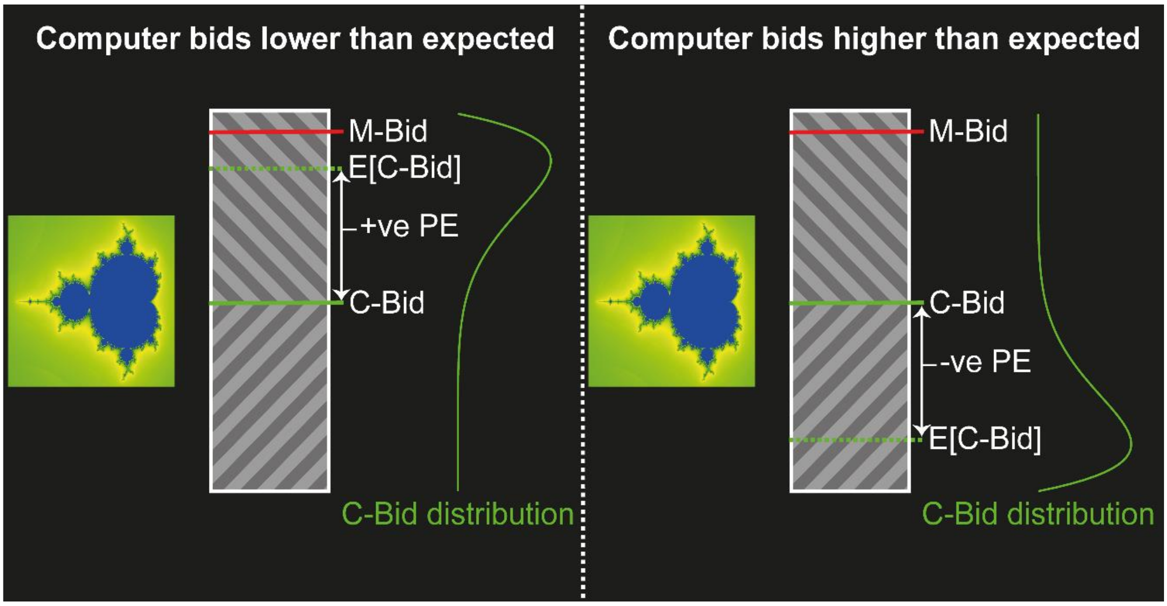


Fig. 6.4) The computer-bid distribution could be manipulated to investigate PEs in response to costs in a quantitative manner. On both the left and right the realised monkey-bid (M-Bid) and computer-bid (C-Bid) are the same, and so the outcome of the trial will be the same in terms of reward and budget juice delivery, however, the PE is dependent upon the distribution of computer bids, allowing the expected computer-bid (E(C-Bid)) to be manipulated independently.

It may also be possible to use such a paradigm to find out whether the subject considers the costs for the different reward and budget juices separately, or whether the LHb codes an integrated overall PE expressed in terms of the total costs experienced in the trial. Thus, we can investigate whether LHb neurons code the PE in terms of the difference between overall expected value and the realised value, or, whether they reflect changing predictions relating to specific stimuli and rewards at specific time points.

Appendices

Appendix 1 – BDM theory

Proofs and descriptions of the optimal BDM strategy and the expected costs of deviation from that optimality are provided here for reference. The axioms of the Von-Neumman-Morgenstern (VNM) theorem and the expected utility (EU) theory on which it is based are also briefly outlined to support discussion of the optimal strategy in the BDM as well as some of the theoretical challenges to the incentive compatibility of the BDM mechanism.

Expected Utility (EU) theory and the VNM axioms:

The expected utility (EU) hypothesis holds that the utility of a lottery is the statistical expectation – i.e. a probability-weighted sum – of the subject's utilities for the different outcomes in that lottery. Where ***L*** is a lottery with outcomes ***A*** and ***B*** occurring with probability ***p*** and ***1-p*** respectively, and *u* is a utility function describing the subject's preferences, EU theory holds that the utility for that lottery is given by:

$$\boldsymbol{u}(\boldsymbol{L}) = \boldsymbol{p}\boldsymbol{u}(\boldsymbol{A}) + (\mathbf{1} - \boldsymbol{p})\boldsymbol{u}(\boldsymbol{B})$$

The EU theory may not provide a perfect description of economic behaviour under risk, and fails to account for observed deviations from its predictions in some experimental settings (for example, it cannot explain loss aversion). Nevertheless, the theory provides a good first approximation to human and animal decision-making under probabilistic conditions, and, most importantly, has provided a good predictive model in several studies of both human and monkey economic behaviour.

The VNM utility theorem proposes four axioms under which it can be shown that a subject faced with different options, each with probabilistic, or 'risky', outcomes, should behave as if they are maximising the value of a utility function defined over all the potential outcomes. In other words, the VNM utility theorem outlines the necessary and sufficient conditions under which the expected utility (EU) hypothesis holds.

The four fundamental VNM axioms are as follows:

1. **Completeness –** For any lotteries, {***A, B***}, exactly one of the following relations must hold true:

$$\boldsymbol{A} \succ \mathbf{B},\ \mathbf{B} \succ \mathbf{A},\ \mathbf{A} \sim \mathbf{B}$$

Where '$\succ$'describes a preference and '$\sim$'describes indifference – it is also possible to have a weak preference, whereby something is indifferent or preferred, '$\succsim$'.

Essentially, the completeness axiom states that the subject's preferences are defined for all lotteries.

2. **Transitivity –** Preferences are consistent across any number of options, for example, for three lotteries {***A, B, C***}:

If $\boldsymbol{A} \succsim \mathbf{B}$ and $\boldsymbol{B} \succsim \boldsymbol{C}$, then $\boldsymbol{A} \succsim \boldsymbol{C}$

3. **Continuity –** This axiom assumes that there exists a lottery between a preferred and less preferred option for which the subject is indifferent to a middle option. Thus, there preferences are described smoothly between 'less preferred' and 'more preferred'.

If $\boldsymbol{A} \succsim \mathbf{B} \succsim \boldsymbol{C}$, then there exists some probability, $\boldsymbol{p}$, such that:

$$\boldsymbol{pA} + (\mathbf{1} - \boldsymbol{p})\boldsymbol{C} \sim \boldsymbol{B}$$

4. **Independence –** The independence of irrelevant alternatives states that preferences hold, regardless of other additional outcomes:

If $\boldsymbol{A} \succsim \mathbf{B}$, then for any other reward $\boldsymbol{Y}$, and probability, $\boldsymbol{p}$

$$\boldsymbol{pA} + (\mathbf{1} - \boldsymbol{p})\boldsymbol{Y} \succsim \ \boldsymbol{pB} + (\mathbf{1} - \boldsymbol{p})\boldsymbol{Y}$$

This axiom implies another axiom of importance to the BDM mechanism when the item being bid for is a lottery: the reduction of compound lotteries.

Reduction of compound lotteries:

The independence axiom implies that given outcomes {***A**, **B***}, and any probabilities, {***p**, **q**, **r***}, such that $\boldsymbol{rq = p}$:

For any lottery $\boldsymbol{L}$, such that $\boldsymbol{L = qA + (1-q)B}$

$$\boldsymbol{pA + (1-p)B \sim rL + (1-r)B}$$

This axiom effectively states that the subject is indifferent between the compound lottery and an equivalent simple lottery – that is, the lottery in which the probabilities of lottery outcomes have been multiplied out. For the BDM for risky goods to be incentive compatible, this reduction must be possible; it must be possible to weigh lotteries by their probability of occurring, and therefore multiply out the outcomes of that lottery by the probability of that lottery being realised.

Consider a BDM for a lottery, $\boldsymbol{L}$**,** in which the total budget is given by $\boldsymbol{T}$, the bid is given by $\boldsymbol{B}$, and the probability of winning is given by $\boldsymbol{p_w}$. Then the following holds true if the subject's bid is the maximum possible amount of the budget that the subject is willing to sacrifice to gain the lottery:

$$\boldsymbol{T \sim p_w(L + T - B) + \ (1 - p_w)T}$$

$$\boldsymbol{T \sim p_w L - \ p_w B + \ T}$$

$$\boldsymbol{T \sim p_w(L - B) + \ T}$$

The above equality is clearly satisfied when $\boldsymbol{L = B}$, that is, when the utility of the bid is equal to the utility of the lottery – i.e. its certainty equivalent. However, for this to hold true, we must allow for the substitution of lotteries, and, the linear weighting of probabilities in the lottery $\boldsymbol{L}$ by $\boldsymbol{p_w}$. Moreover, it is also apparent from the above, that this indifference between the budget and the bid is satisfied when $\boldsymbol{L = B}$, regardless of the probability of winning the auction – again, this only holds if the independence axiom is satisfied.

In short, the VNM theorem implies that the subject's overall expected utility in the BDM should reflect the utilities of the outcomes after all probabilities of lottery outcomes are multiplied out - thus, the certainty equivalent measured in the compound lottery of the BDM reflects the subject's value for the lottery within it.

Optimal BDM strategy:

The optimal strategy in the BDM is the same as that in a second-price sealed-bid, or Vickrey, auction. Here, we present the optimal strategy for a second-price sealed-bid auction, as adapted from Milgrom and Weber's (1982) more comprehensive proof[96].

To find the optimal strategy for bidder $\boldsymbol{i}$, assuming they have a smooth, continuous and differentiable utility function increasing in income, $\boldsymbol{U_i}$, let $\boldsymbol{v_i}$ represent the value placed on the good by bidder $\boldsymbol{i}$, who places a bid, $\boldsymbol{b_i}$, to obtain the good against other bidders. If bidder $\boldsymbol{i}$ wins the auction, they will derive utility from the difference between the second highest bid - the price, $\boldsymbol{p}$ - and their valuation; this is given by $\boldsymbol{U_i(v_i - p)}$. If bidder $\boldsymbol{i}$ loses, their monetary value from participation is taken as zero. At the time of bidding, the price, $\boldsymbol{p}$, is effectively a random variable. Suppose that bidder $\boldsymbol{i}$ has an expectation of the price characterised by the cumulative distribution function $\boldsymbol{F_i(p)}$, with support $[\underline{\boldsymbol{p_i}}, \overline{\boldsymbol{p_i}}]$ and probability density function $\boldsymbol{f_i(p)}$. Expected utility ($\boldsymbol{E[U_i]}$) is therefore expressed by the following equation:

$$\boldsymbol{E[U_i]} = \int_{\underline{\boldsymbol{p_i}}}^{\boldsymbol{b_i}} \boldsymbol{U_i(v_i - p)\, dF_i(p)} + \int_{\boldsymbol{b_i}}^{\overline{\boldsymbol{p_i}}} \boldsymbol{U_i(0)}$$

$$= \int_{\underline{\boldsymbol{p_i}}}^{\boldsymbol{b_i}} \boldsymbol{U_i(v_i - p) f_i(p) dp} + \int_{\boldsymbol{b_i}}^{\overline{\boldsymbol{p_i}}} \boldsymbol{U_i(0)}$$

We can normalise the utility of zero money to zero, such that $\boldsymbol{U(0) = 0}$:

$$\boldsymbol{E[U_i]} = \int_{\underline{\boldsymbol{p_i}}}^{\boldsymbol{b_i}} \boldsymbol{U_i(v_i - p) f_i(p) dp}$$

The maximum of this function is found when it's first derivative with respect to the bid, $\boldsymbol{b_i}$, is set equal to zero:

$$\frac{\boldsymbol{\partial E[U_i]}}{\boldsymbol{\partial b_i}} = \boldsymbol{U_i(v_i - b_i) f_i(b_i) = 0}$$

It is trivially apparent that this equation is satisfied when $\boldsymbol{b_i = v_i}$, i.e. when player $\boldsymbol{i}$'s bid is set equal to their value.

Expected cost of misbehaviour (ECM):

Calculation of the ECM depends upon finding difference in the expected profits of each possible bid and that of the optimal bid. In this case expected profit is simply the difference in subjective value (measured in units of the budget) between the expected overall payoff at the end of the trial and the starting budget.

First, note that the expected profit ($\boldsymbol{E}[\boldsymbol{\pi_i}]$) in the BDM, for subject $\boldsymbol{i}$, submitting a bid, $\boldsymbol{b_i}$ and having a value, $\boldsymbol{v_i}$, against a computer drawing a price, $\boldsymbol{p}$, from a probability density function, $\boldsymbol{f}(\boldsymbol{p})$, with cumulative distribution function, $\boldsymbol{F}(\boldsymbol{p})$ - such that the probability of winning is $\boldsymbol{F}(\boldsymbol{b_i})$ - is given by[97]:

$$\boldsymbol{E}[\boldsymbol{\pi_i}] = [\boldsymbol{v_i} - \int_{-\infty}^{\boldsymbol{b_i}} \left(\frac{\boldsymbol{f}(\boldsymbol{p})}{\boldsymbol{F}(\boldsymbol{b_i})}\right) \boldsymbol{p}\, \boldsymbol{dp}](\boldsymbol{F}(\boldsymbol{b_i}))$$

The expected cost of misbehaviour (ECM) is simply the difference between the expected profit of the optimal bid, where $\boldsymbol{v_i} = \boldsymbol{b_i}$, and the expected profit of some other sub-optimal bid, where $\boldsymbol{v_i} \neq \boldsymbol{b_i}$:

$$\boldsymbol{ECM_i}(\boldsymbol{v_i}, \boldsymbol{b_i}) = \boldsymbol{E}[\boldsymbol{\pi_i} | \boldsymbol{v_i} = \boldsymbol{b_i}] - \boldsymbol{E}[\boldsymbol{\pi_i} | \boldsymbol{v_i} \neq \boldsymbol{b_i}]$$

For a BDM task with computer bids drawn from a uniform distribution with support $[\mathbf{0}, \mathbf{1}]$, the profit function is given by:

$$\boldsymbol{E}[\boldsymbol{\pi_i}] = \left(\boldsymbol{v_i} - \frac{\boldsymbol{b_i}}{\mathbf{2}}\right) \boldsymbol{b_i}$$

Therefore, the ECM is:

$$\boldsymbol{ECM_i} = \frac{(\boldsymbol{v_i} - \boldsymbol{b_i})^{\mathbf{2}}}{\mathbf{2}}$$

We can express this in terms of a parameter describing the degree of deviation from the subject's true value, $\boldsymbol{\gamma_i}$, such that $\boldsymbol{\gamma_i} = \boldsymbol{v_i} - \boldsymbol{b_i}$:

$$\boldsymbol{ECM_i} = \frac{{\boldsymbol{\gamma_i}}^{\mathbf{2}}}{\mathbf{2}}$$

It is clear from this that the magnitude of the subject's value has no effect on their incentives for truthful bidding in the BDM, and is a function of the disparity between the subject's bid and their true value only. Therefore, incentives to bid truthfully in the

BDM should not vary based on the value of the reward bit the costs of a given deviation from optimality are effected by the computer-bid distribution that the subject faces.

While a closed form solution for the ECM with a Beta distribution of computer bids does not exist, simulation of expected payoffs can provide approximate measures. Lusk et al. (2007) conducted 20 000 such simulations for each of several auction computer-bid parameters. Similar simulations were used when presenting ECM in this thesis (Appendix 4).

Representing ECM in terms of utility:

It is worth noting that the ECM describes costs in terms of the budget that is used to bid for the reward, that is, it is a cost in terms of the numeraire, and reflects subjective value, but not necessarily the utility - though it is related to the utility by a monotonic function.

Presented here is a simple extension to the work of Lusk et al. (2007) in considering how representing the ECM in terms of utility could lead to different expected rates of learning when compared to a representation of the ECM in terms of subjective value. If reinforcement signals such as the prediction error (PE) are expressed in terms of the utility[3], rather than in terms of subjective value – or units of the numeraire – then the gradient of payoffs in the BDM (and therefore the gradient of ECMs) will differ to that expressed in terms of the numeraire in a manner that is dependent upon the shape of the subject's utility function.

Consider the following simulated bidder facing a uniform computer-bid distribution with support **[0, 1]** over some numeraire reward, ***R***, and bidding for a reward with a subjective value of 0.5 units of ***R***.

Three different hypothetical utility functions are shown: concave ($\boldsymbol{u}(\boldsymbol{R}) = \boldsymbol{R}^{\boldsymbol{1/2}}$), convex ($\boldsymbol{u}(\boldsymbol{R}) = \boldsymbol{R}^{\boldsymbol{2}}$), and a linear function ($\boldsymbol{u}(\boldsymbol{R}) = \boldsymbol{R}$), wherein the utility faithfully reflected the objective magnitudes of the numeraire (fig. A1.1a). By representing the payoffs on simulated trials in terms of these utilities, rather than in units of the numeraire, expected profits could also come to be expressed in terms of utility, leading to a different gradient of ECMs for each of the utility functions (fig. A1.1b).

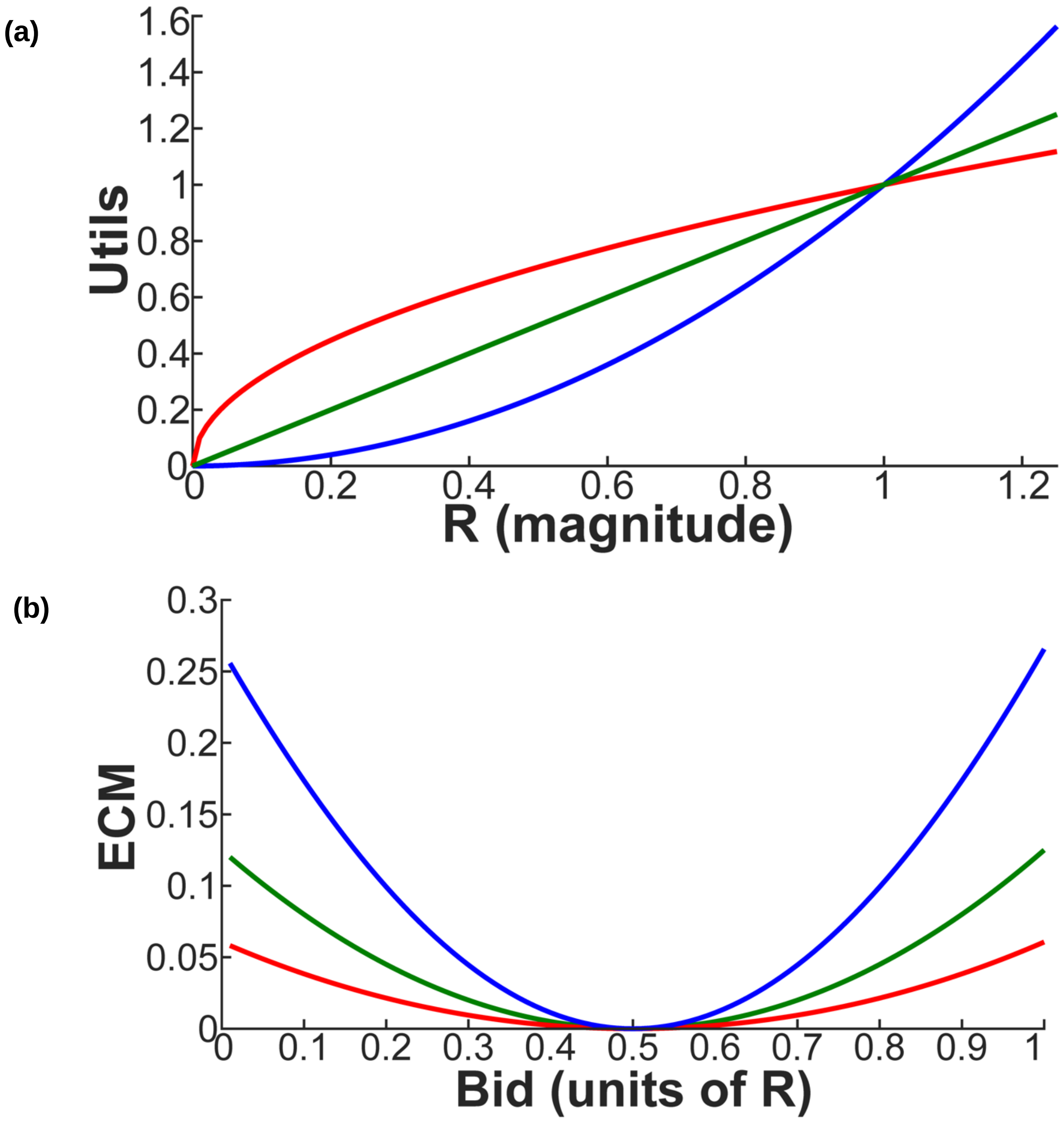


Fig. A1.1a) Convex (blue), concave (red), and linear (green) utility functions were used to simulate the expected payoffs of each bid for a simulated subject bidding for an object with a subjective value of 0.5 units of ***R***. **b)** The ECM was calculated using payoffs expressed in terms of their utils, however, the ECM derived from the linear utility function (green) is equivalent to an ECM calculated in units of the numeraire, ***R***. Compared to the ECM as expressed in terms of the numeraire (green), there is a steeper gradient of costs for a subject with a convex utility function (blue) and a shallower gradient of costs for a subject with a concave utility function (red).

If the reinforcement signal is expressed in terms of utility, rather than units of the numeraire, then the steeper payoff, and therefore cost, gradient for a subject with a convex utility function should drive faster learning of the optimal strategy, especially if the subject must search the bidding-range to find the optimal action. Moreover, the steeper cost gradient of the subject with a convex utility function should also provide a greater incentive for truthful bidding.

Manipulations of the identity of the numeraire could be used to test BDM performance after the task has been learnt. Two different sets of budget and reward, one with an overall convex and the other with a concave utility function, could be used to test the hypothesis that goods over which the utility function is convex lead to greater incentives to place accurate bids – i.e. the variance in the monkey's bids should be lower, and the mean bid should be closer to optimal when the monkey's utility function is convex.

Appendix 2 – Experimental methods

Animals and experimental setup:

Two purpose-bred (at the MRC centre for macaques) and group-housed male rhesus monkeys (Macaca mulatta) were used for this study (weighing 7.9 kg and 10.8 kg) over 24 months and 36 months of daily testing (1-2 hours each day), respectively. The Home Office of the United Kingdom approved all experimental protocols and procedures which were all in line with the animals scientific procedures act (ASPA) and also approved by the local Animal Welfare and Ethics Board (AWERB). During experimental sessions animals sat in a primate chair (Crist Instruments) positioned 60cm from a computer monitor. A joystick (Biotronix Workshop) allowing both up-down and left-right movements was used by the animals to report choices in the BDM and BC tasks, respectively.

Joystick position and touch-key data and digital task event signals were sampled at 2kHz, and stored at 200 Hz (joystick) or 1 kHz (task events). Liquid reward was delivered by a computer-controlled solenoid liquid valve (~0.006ml/ms opening time). The behavioural tasks were controlled by custom-made software (MATLAB; The MathWorks) running in conjunction with the Psychophysics toolbox[98] on a Microsoft Windows 7 computer.

Joystick control:

The joystick was used to move an on-screen marker in both the BDM and BC tasks. Voltage outputs for horizontal and vertical movement of the joystick were separated, and for both, at its central position, the joystick voltage output was 0v. A maximal forward or rightward movement produced a voltage of 5v, and a maximal backward or leftward movement produced a voltage of -5v. The position of the on-screen marker was modulated by the following equation, where $\boldsymbol{G}$ is the gain or amplification applied to the voltage modulation, $\boldsymbol{V}$, and $\boldsymbol{P}$ is the pixel position of the centre of the on-screen marker at time, $\boldsymbol{T}$:

$$\Delta_{\boldsymbol{T}} = \boldsymbol{GV}$$

$$\boldsymbol{P_T} = \boldsymbol{P_{T-1}} + \Delta_{\boldsymbol{T}}$$

Thus, the value of $\boldsymbol{P}$ changes more quickly with larger deflections of the joystick in the direction of its movement. In the BDM, forward and backward deflections of the joystick move the bid-marker up and down the budget-bar respectively, and the value of $\boldsymbol{P}$ is limited such that its maximum and minimum values are set to the maximum and minimum pixel positions of the budget-bar. Moreover, in the BDM the value of $\boldsymbol{G}$ was the same for movements in both directions.

In the BC task, the value of $\boldsymbol{G}$ was set to depend on whether $\boldsymbol{V}$ took a positive or negative value, thus the gain could be set to have different values for rightward and leftward movements of the joystick. This was used to counteract the effects of side-bias on the monkey's choices in the BC task (Ch. 5.1), with values of $\boldsymbol{G}$ being chosen for each direction such that the monkey made choices without a statistically significant side-bias where both the left and right-hand-side offers were the same.

A minimum threshold of 2% of this maximal voltage displacement was applied in every direction, such that any voltage with absolute magnitude of 0.1v was treated as a 0v modulation, and therefore did not produce any deflection of the on-screen markers. This gave a centre threshold within which the monkey could make slight deviations of the joystick in each direction without experiencing a 'not-centred' error – the monkeys found it extremely difficult to hold the joystick perfectly still in the central position, and therefore such a tolerance window was necessary to allow them to participate in the task without becoming too frustrated.

Moreover, without such a window, small movements of the joystick would be detected and lead to erratic deflections of the on-screen markers, preventing the monkeys from properly stabilising the joystick position.

After the introduction of the BC task (Ch. 5.1), the joystick had to be stabilised for 250ms (500 samples at 2kHz) such that the value of $\boldsymbol{P}$ was the same for all samples. Failure to do so resulted in a 'no-choice' error. At this point we also introduced a 'not-centred' error, where any deviations beyond the aforementioned central tolerance window outside the 'choice' epoch led to a 'not-centred' error.

Errors led to a blue time-out screen that was the same duration as the remaining trial time plus 3s.

Touch-key:

The joystick had a built-in touch-key which could detect skin contact from the monkeys. Any release of the joystick at an inappropriate time would be detected and result in a ‘no-hold’ error. The monkey had to maintain hold of the joystick in the ‘offer’ and ‘choice’ epochs of both the BDM and BC tasks.

Solenoid reward-delivery:

To deliver juice-reward and water-budget in our tasks we used a solenoid delivery system with voltage pulses controlling the opening time of the solenoid. There was an approximately linear relationship between solenoid opening time and the volume of water/juice delivered, and we tested and calibrated the opening times so that we could deliver the appropriate volumes of the different rewards in the task.

The measurements during calibration of our solenoid juice-delivery system showed a mean standard deviation of 0.06ml at a given opening time. We did not test the monkey’s ability to discriminate different volumes of juice, except through their performance in the BC task - which indicated an ability to discriminate differences of ~0.1ml in different versions of the task.

This degree of variability in the volume of reward delivered at a given solenoid opening time could limit the monkey’s ability to discriminate the small differences in expected payoffs that result from different bids in the BDM (Ch. 2.1), as these variations in delivered volume may be indistinguishable from the variability of the solenoid itself.

Increasing the water-budget volume and the volume of the juice-rewards reduces the relative magnitude of the solenoid’s variable reward delivery, as the standard deviation of the delivered volume was the same regardless of the mean volume delivered.

With a larger water-budget volume, the ECM will be greater for the same pixel distance displacement of the bid-marker from the optimal bid, and the relative contribution of variability in the solenoid delivery will be reduced. Thus, the monkey should be able to discriminate differences in expected payoff at smaller relative distances between the actual and optimal bid.

Appendix 3 – Statistical methods

All statistical tests were run using MATLAB (The Mathworks) software[99], the specific functions that were used are referred to as they become relevant. An alpha (Type I; false-positive) level of 0.05 was used throughout this thesis whenever groups were being compared. If multiple comparisons were conducted, then we always applied a Bonferroni correction to the value of alpha.

Logistic regression:

Binary choice (BC) tasks (Ch. 2.3) were used to infer the water-budget values of different volumes of juice-reward. Two different versions of this task were used in this thesis. First, we trained monkeys in a simple BC task in which option ***A*** was a water-budget offer and option ***B*** was a specific juice-reward offer (Ch. 5.1). However, the main task that we used was a binary-choice-bundle (BCb) task, in which option ***A*** was a bundle of juice-reward and water-budget and option ***B*** was always the total water-budget volume. We fit a logistic regression model[100] to the monkey's choice data for both the BC and BCb tasks to find the monkey's value for the juice-reward in terms of water-budget.

The BCb task should provide a better analogue for the BDM due to its assessment of the monkey's value for the juice-reward at the same wealth level as in the BDM (Ch. 5.3), and, as the BCb is the main binary choice task that was used in this thesis we describe logistic regression primarily in terms of the BCb task[*], where ***X*** is the volume of water-budget offered in ***A***:

$$\boldsymbol{y = \beta_0 + \beta_1(X)}$$

$$\boldsymbol{y = \begin{cases} 1 \; option\ A\ chosen \\ 0 \; option\ B\ chosen \end{cases}}$$

$$\boldsymbol{Proportion\ of\ A\ choices = \frac{1}{1 + e^{-y}}} \qquad [\boldsymbol{A3.1}]$$

[*] The method is the same for the simple BC task, and any differences are described where they are of relevance.

The proportion of ***A*** choices is obtained for each value of ***X*** that is tested in the BCb task. We then fit the logistic function of the form shown in equation A3.1 to these choice data, where $\boldsymbol{\beta_0}$ is a measure of choice bias and $\boldsymbol{\beta_1}$ is a measure of the monkey's sensitivity to increases in the volume of water-budget offered in the bundle. After values for these two coefficients were estimated, we could use these to find the volume of budget-juice in the bundle, $\mathbf{X_{50}}$, for which the subject is indifferent between options ***A*** and ***B*** – i.e. where the proportion of ***A*** choices is equal to 1/2. We could use the estimated coefficients to find the value of $\mathbf{X_{50}}$, first, by substituting the proportion of choice of ***A*** at indifference in equation A3.1:

$$\frac{1}{2} = \frac{1}{1 + e^{-y}}$$

$$2 = 1 + e^{-y}$$

$$e^{-y} = 1$$

$$-y = 0$$

$$-\beta_0 - \beta_1(X_{50}) = 0$$

$$X_{50} = -\frac{\beta_0}{\beta_1} \qquad [A3.2]$$

If this was the simple version of the BC task, in which ***A*** was constituted of water-budget offers and ***B*** of a given volume of juice-reward, then $\mathbf{X_{50}}$ in equation A3.2 would provide the value of ***B*** in units of water-budget. However, in the case of the BCb task, $\mathbf{X_{50}}$ is the amount of water-budget that must be offered alongside the juice-reward for the monkey to be indifferent between ***A*** and ***B***. That is, the difference between the total budget volume in ***B*** and the amount of water-budget offered in ***A*** is compensated for by the presence of the reward-juice in ***A***. Therefore, an equivalent to the BDM bid, or, the value of the reward-juice, should be given by:

$$BCb\ reward\ juice\ value = Maximum\ budget\ volume\ -\ X_{50}$$

Logistic regression was conducted using the '*glmfit*' function in MATLAB, with a binomial distribution and a logit link function. We also acquired upper and lower 95% confidence interval fits for the logistic function and used these to acquire a range of values within which the true value is predicted to lie with 95% confidence. We used this range when comparing mean bids in the BDM and values inferred using logistic regression in the BCb task. Moreover, this allowed us to establish the precision of value estimation in the BCb task, which was also to be compared with the precision of the BDM method.

Binomial test:

We assessed whether monkeys were showing a side bias during their performance of the binary choice tasks. First, we would check whether there was an imbalance in the proportion of right-hand-side (RHS) or left-hand-side (LHS) choices over all trials at the end of the session – as the side of presentation of different offers was randomised there shouldn't have been a side bias over the many trials tested. Second, if a significant bias was detected, we would use a version of the task that offered the same reward on either side – in this case there should be no difference in the proportion of LHS and RHS choices, and any statistically significant preference would reveal the presence of a side bias.

To identify whether such biases existed, we used the '*binofit*' function in MATLAB, and used this to acquire a 95% confidence interval for the number of LHS choices that an unbiased decision-maker would make given the number of trials that were tested. If the monkey made more/fewer LHS choices than this, then we identified this as evidence of the presence of a LHS/RHS bias respectively.

If such a bias was identified in the main binary choice task then we would conduct a test with same rewards on either side, as described above, and perform the same statistical testing on the LHS choices in that session. We used this to titrate the gains of the LHS and RHS deflections of the joystick such that there was no longer a detectable side bias in the monkey's choices (Appendix 2).

Spearman's rank correlation:

Spearman's rank correlation is a nonparametric measure of the dependence between the rankings of two variables. Thus, the absolute values do not influence the strength of the Spearman's rank correlation between two variables - only their relative rankings do. Therefore, Spearman's rank correlation describes how well the relationship between the two variables being correlated can be described using a monotonic function, without assuming linearity.

For this reason, Spearman's Rho, the measure of the strength of the correlation, is used throughout this thesis to summarise the relationship between the monkey's bids and the reward magnitudes. Because the relationship between reward magnitudes and their subjective values is often non-linear, Spearman's Rho provides a suitable metric of consistency.

Moreover, the most basic test of performance in the BDM is whether or not the values inferred from the bids for different rewards reflects their ordinal ranking. That is, does the monkey bid more for more preferred rewards? This is entirely captured by Spearman's Rho, with perfectly consistent bidding having a Spearman's Rho that approaches 1.

To implement Spearman's rank correlations, we used MATLAB's '*corr*' function. In cases where the partial correlation coefficients were desired – an analysis that controls for the influence of other factors - MATLAB's '*partialcorr*' function was used.

Fisher transformation of Spearman's Rho:

In measuring the monkeys' performance, we have usually attempted to find the mean and standard deviation of Spearman's Rho for the correlation between the monkey's bids and the reward magnitudes across several sessions of the same type. Where the presence of strongly outlying sessions has reduced the certainty in these estimates, we have sought to assess the relationship over the population of all bids and rewards in those sets of sessions.

This gives a Spearman's Rho for the population data, and 95% confidence intervals can be derived through a Fisher transformation of this Spearman's Rho – allowing for comparison of two populations for which the population data, rather than session-by-session values of Spearman's Rho, have been used.

The Fisher transformation to obtain the upper and lower confidence intervals (CI) for Spearman's Rho proceeds as follows, where ***n*** is the number of observations used to infer the value of Rho:

$$\boldsymbol{Delta} = \mathbf{1.96}\left(\frac{\mathbf{1}}{\sqrt{\boldsymbol{n}-\mathbf{3}}}\right)$$

$$\boldsymbol{Upper\ CI} = \mathbf{tanh}(\mathbf{arctanh}(\boldsymbol{Rho}) + \boldsymbol{Delta})$$

$$\boldsymbol{Lower\ CI} = \mathbf{tanh}(\mathbf{arctanh}(\boldsymbol{Rho}) - \boldsymbol{Delta})$$

Identification of outliers:

Given the exploratory nature of the experiments in this thesis and the limited number of sessions for the various experimental conditions, we avoided excluding any sessions as outliers. However, where we have considered the possibility that sessions may be unrepresentative we have analysed the data following exclusion of data-points that are more than 3 scaled median absolute deviations (MAD) away from the median. Conclusions drawn from the data never depended upon analyses that excluded such 'outliers', but those analyses did allow us to further interrogate limited and noisy data-sets.

Scaled MAD is calculated by taking the median of the absolute difference between each element, $\boldsymbol{x_i}$, in a sample, ***S***, and the median of that sample, and multiplying this difference by a scaling factor, such that:

$$\mathbf{Scaled\ MAD}(\boldsymbol{S}) = \mathbf{1.4826}\ \boldsymbol{x}\ \boldsymbol{median}(|\boldsymbol{x_i} - \boldsymbol{median}(\boldsymbol{S})|)$$

<u>Analysis of variance (ANOVA):</u>

One-way Analysis of Variance (1-way ANOVA) was often used in conjunction with multiple comparisons tests to identify whether the monkey's bids could be used to differentiate between different reward magnitudes, and, whether the bids for different reward magnitudes were significantly different to one another.

While Spearman's rank correlation gave a good measure of the consistency of bidding and its ordinality, it did not tell us which groups of bids (for different rewards) were significantly different. It is possible to have a reasonably high value of Spearman's Rho without being able to statistically distinguish the mean bids for all the rewards that were tested.

Therefore, the omnibus 1-way ANOVA test was used to determine whether reward magnitudes influenced the monkey's bidding, with follow-up Bonferroni-corrected multiple-comparisons tests being used to identify those groups of rewards for which the bids were significantly different.

We used MATLAB's '*anova1*' function for 1-way ANOVAs and the '*anova2*' function for 2-way ANOVAs.

Effect sizes were captured by the eta-squared, or $\boldsymbol{\eta^2}$, parameter. This was calculated by dividing the sum of squares due to differences between groups in a factor by the sum of squares total:

$$\boldsymbol{\eta^2_{Factor}} = \frac{\boldsymbol{SS_{Factor}}}{\boldsymbol{SS_{Total}}}$$

<u>Two-sample t-tests:</u>

Two-sample t-tests were used to compare the means of different groups of data throughout this thesis. The test made use of MATLAB's '*ttest2*' function, and variances were not assumed to be equal unless stated otherwise – this is called Welch's t-test and only assumes normality.

Welch's t-test performs similarly to student's t-test when variances and sample-sizes of the two groups being compared are equal, but is more robust than student's t-test when group sizes and/or variances are unequal. Due to the robust nature of Welch's t-test, there is little advantage to pre-testing the two samples for equality of variances[101]. Therefore, for simplicity, Welch's t-test is used for all two-sample t-tests in this thesis.

Appendix 4 – Simulations

Simulations of the BDM and BC tasks:

We simulated two types of decision-makers in both the BDM and BC tasks, either a decision-maker who made theoretically optimal decisions without noise, or, a decision-maker who made decisions according to the best strategy, but with noise around the optimum.

Simulations of the BDM, BC and BCb tasks were primarily used to assess differences in average per-trial payoff across the three task types (Ch. 5.3, fig. 5.12).

In the case of the BDM task, we fit a truncated normal distribution to Ulysses' bids in a relevant set of sessions for a specific reward magnitude. We chose a set of sessions that matched key simulated task parameters, such as the water-budget volume. We used MATLAB's '*fitdist*' function to generate a normal distribution* that was fit to the monkey's bids for that reward, we then used the '*truncate*' function to truncate this distribution to the limits of possible bids, i.e. between 0 and the maximum budget volume.

This distribution could be used to generate bids like the monkey in a simulated BDM task: on each simulated trial a BDM bid would be generated from this distribution as well as a random computer-bid, with the outcomes of the trial being determined according to the BDM mechanism. Thus, the average per-trial payoff could be calculated for this simulated bidder. The same procedure was followed for a simulated optimal bidder, but in this case the simulated bidder always bid the exact same amount: their value for the reward.

For the simulated BC and BCb tasks, we fit a logistic function to the monkey's choice data. This gave a value for β_1, a measure of noise in the monkey's decision-making process. Using the transformation shown in equation A3.3 (Appendix 3), we could find an equivalent normal standard deviation and used this to generate noise on each trial. This noise was applied to the simulated subject's underlying value for the reward, and, as decisions were simulated after the application of this noise, the

* We also tested a beta distribution fit to the monkey's bids and found no significant difference in the results thus generated.

monkey sometimes made erroneous decisions given their true underlying value for the reward. In the case of the optimal decision maker in the BC and BCb tasks, no noise was applied to the value of the reward, and their decisions therefore had the form of a step function with a step at the simulated subject's reward value.

In addition to this, in Chapter 1 (fig. 1.2) we use a simulated decision-maker in a binary-choice task to present the potential effects of sensory specific satiety on choice behaviour in the BC task. To do this, we used the same noisy decision-maker, but also applied a change in their value for option ***A***, $\boldsymbol{AV}$, to reflect the effects of sensory specific satiety for the two different rewards. Consuming more of ***A*** should lead to a relative reduction in the value of ***A*** compared to ***B***, while consuming more of ***B*** should lead to a relative increase in the value of ***A*** in terms of ***B***. This modulation in the value was mediated by the product of a constant and the volume consumed:

$$\boldsymbol{AV} = \boldsymbol{AV} - (\mathbf{0.025} \times \boldsymbol{VolA}) + (\mathbf{0.025} \times \boldsymbol{VolB})$$

Simulations for calculation of the ECM

While a closed-form solution for the ECM in the BDM with a uniform distribution of computer bids does exist, there is no such solution for a bidder facing a Beta distribution of computer bids. As we used the Beta (4,4) distribution for most of our experiments – due to an observed tendency of both monkeys to place bids in the centre of the bidding range during training of the task – we had to simulate the payoffs of different bids when calculating the ECM.

In replication of Lusk et al. (2007), to calculate the ECM we specified the reward value and used 20 000 simulated trials at each bid between the minimum and maximum bids in steps of 1% of the total bidding range. So, if the water-budget was 1.2ml, we simulated 20 000 trials for each bid between 0ml and 1.2ml, in steps of 0.012ml. The simulated bidder made the same bid at each of these steps and the computer-bid was drawn randomly from the relevant Beta distribution on each trial. We then found the mean payoff of each bid in terms of the water-budget by averaging across these 20 000 repetitions. The ECM was then calculated by finding the difference between the mean simulated payoff for each bid and the mean simulated payoff of the optimal bid.

The method described above gave us a measure of the ECM across the range of possible bids in a task given the monkey's value for the reward and the computer-bid distribution that they faced. However, we also wanted to know what the realised difference in payoffs was for each monkey bid compared to an optimal bidder. For this, we simply simulated an optimal bidder facing the computer bid that the monkey experienced on each trial. This gave us a measure of how much better the monkey would have done by bidding optimally given the computer bids that were realised in the task, and we used this to calculate overall and per-trial differences in payoffs.

References

[1] Becker, G. M., DeGroot, M. H., Marschak, J. (1964) “Measuring utility by a single-response sequential method”. *Behavioural Science* 9(3): 226-32.

[2] Mas-Colell, A., Whinston, M. D., Green, J. R. (1995) “Microeconomic theory”. *Oxford University Press* p. 5.

[3] Stauffer, W. R, Lak, A., Schultz, W. (2014) “Dopamine reward prediction error responses reflect marginal utility”. *Current Biology* 24: 2491-2500.

[4] Samuelson, P. (1937) “A note on the measurement of utility”. *The Review of Economic Studies* 4: 155-161.

[5] Fredercik, S., Loewenstein, G., O’Donoghue, T. (2002) “Time discounting and time preference: a critical review”. *Journal of Economic Literature* 40: 351-401.

[6] Bernoulli, D. (1954); Latin original (1738) “Exposition of a new theory on the measurement of risk”. *Econometrica* 22: 23–36.

[7] Tobler, P. N., Fiorillo, C. D., Schultz, W. (2005) “Adaptive coding of reward value by dopamine neurons”. *Science* 307: 1642-1645.

[8] Hurwicz, L. (1960) “Optimality and Informational Efficiency in Resource Allocation Processes”. *Mathematical Methods in the Social Sciences*, edited by Arrow, Karlin and Suppes, Stanford University Press.

[9] Kahneman, D., Knetsch, J. L., Thaler, R. H. (1990) “Experimental Tests of the Endowment Effect and the Coase Theorem”. *Journal of Political Economy* 98: 1325-1348.

[10] Lusk, J. L., Fox, J. A. (2003) “Value Elicitation in Laboratory and Retail Environments”. *Economics Letters* 79: 27-34.

[11] Grether, D. M., Plott, C.R. (1979) “Economic Theory of Choice and the Preference Reversal Phenomenon”. *The American Economic Review* 69(4): 623-638.

[12] Lusk, J. L., Fox, J. A., Schroeder, T. C., Mintert, J., Koohmaraie, M. (2001) “In-store Valuation of Steak Tenderness”. *American Journal of Agricultural Economics* 83: 539-550.

[13] Ding, M., Grewal, R., Liechty, J. (2005) “Incentive-Aligned Conjoint Analysis”. *Journal of Marketing Research* 42: 67-83.

[14] Plassmann, H., O’Doherty, J., Rangel, A. (2007) “Orbitofrontal cortex encodes willingness to pay in everyday economic transactions”. *Journal of Neuroscience* 27: 9984-9988.

[15] Weber, B., Aholt, A., Neuhaus, C., Trautner, P., Elger, C. E., Teichert, T. (2007) “Neural evidence for Reference-dependence in real-market-transactions”. *NeuroImage* 35: 441-447.

[16] Chib, V. S., Rangel, A., Shimojo, S., O’Doherty, J. P. (2009) “Evidence for a Common Representation of Decision Values for Dissimilar Goods in Human Ventromedial Prefrontal Cortex”. *The Journal of Neuroscience* 29(39): 12315-20.

[17] Clithero, J. A., McKell Carter, R., Huettel, S. A. (2009) “Local pattern classification differentiates processes of economic valuation”. *NeuroImage* 45: 1329-1338.

[18] De Martino, B., Kumaran, D., Holt, B., Dolan, R. J. (2009) “The Neurobiology of Reference-Dependent Value Computation”. *The Journal of Neuroscience* 29(12): 3833-3842.

[19] Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., Rangel, A. (2008) "Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors". *The Journal of Neuroscience* 28(22): 5623-5630.

[20] Harris, A., Adolphs, R., Camerer, C., Rangel, A. (2011) "Dynamic Construction of Stimulus Values in the Ventromedial Prefrontal Cortex". *PLoS ONE* 6(6): e21074.

[21] Plassmann, H., O'Doherty, J. P., Rangel, A. (2010) "Appetitive and Aversive Goal Values Are Encoded in the Medial Orbitofrontal Cortex at the Time of Decision Making". *The Journal of Neuroscience* 30(32): 10799-10808.

[22] Vickrey, W. (1961) "Counterspeculation, Auctions, and Competitive Sealed Tenders". *The Journal of Finance* 16 (1): 8-37.

[23] Moldovanu, B., Tietzel, M. (1998) "Goethe's Second-Price Auction". *Journal of Political Economy* 106(4): 854-859.

[24] Neumann, J. V., Morgenstern, O. (1953) "Theory of Games and Economic Behavior". *Princeton University Press*, Princeton, NJ.

[25] Krishna, V. (2009) "Auction theory". *Academic Press* p. 3.

[26] Montague, P., Berns, G. (2002) "Neural economics and the biological substrates of valuation". *Neuron* 36: 265-284.

[27] Lak, A., Stauffer, W. R., Schultz, W. (2014) "Dopamine prediction error responses integrate subjective value from different reward dimensions". *PNAS USA* 6: 2343-2348.

[28] Tobler, P. N., O'Doherty, J. P., Dolan, R. J., Schultz, W. (2007) "Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems". *Journal of Neurophysiology* 97(2): 1621-1632.

[29] O'Doherty, J. P. (2004) "Reward representations and reward-related learning in the human brain: insights from neuroimaging". *Current Opinion in Neurobiology* 14(6): 769-76.

[30] Padoa-Schioppa, C., and Assad, J. A. (2006) "Neurons in the orbitofrontal cortex encode economic value". *Nature* 400: 233-238.

[31] Musallam, S., Corneil, B.D., Greger, B., Scherberger, H., Andersen, R. A. (2004) "Cognitive control signals for neural prosthetics". *Science* 305: 258–262.

[32] Mosteller, F., and Nogee, P. (1951) "An experimental measurement of utility". *Journal of Political Economy* 69: 371-404.

[33] Rolls, B. J. (1986) "Sensory-specific satiety". *Nutrition Reviews* 44: 93-101.

[34] Sutton, R. S., Barto, A. G. (1998) "Reinforcement Learning: An Introduction". *Cambridge: MIT Press.*

[35] Katehakis, M. N., Veinott, A. F. (1987) "The Multi-Armed Bandit Problem: Decomposition and Computation". *Mathematics of Operations Research* 12(2): 262–268.

[36] Noonan, M. P., Walton, M. E., Behrens, T. E. J., Sallet, J., Buckley, M. J., Rushworth, M. F. S. (2010) "Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex". *PNAS USA* 107: 20547-20552.

[37] Irwin, J. R., McClelland, G. H., McKee, M., Schulze, W. D., Norden, N. E. (1998) “Payoff Dominance vs. Cognitive Transparency in Decision Making”. *Economic Inquiry* 36: 272-285.

[38] Noussair, C., Robin, S., Ruffieux, B. (2004) “Revealing consumers’ willingness-to-pay: A Comparison of the BDM Mechanism and the Vickrey Auction”. *Journal of Economic Psychology* 25(December): 725-741.

[39] Holt, C. A. (1986) “Preference Reversals and the Independence Axiom”. *American Economic Review* 76: 508-515.

[40] Karni, E., Safra, Z. (1987) “Preference reversals and the observability of preferences by experimental methods”. *Econometrica* 55: 675-685.

[41] Gonzalez, R., Wu, G. (1999) “On the shape of the probability weighting function”. *Cognitive Psychology* 166: 129-166.

[42] Stauffer, W. R., Lak, A., Bossaerts, P., Schultz, W. (2015) “Economic Choices Reveal Probability Distortion in Macaque Monkeys”. *The Journal of Neuroscience* 35(7): 3146-3154.

[43] Kahneman, D., Knetsch, J. L., Thaler, R. H. (1990) “Experimental Tests of the Endowment Effect and the Coase Theorem” *The Journal of Political Economy* 98: 1325-1348.

[44] Thaler, R. H. (1980) “Toward a Positive Theory of Consumer Choice”. *Journal of Economic Behavior and Organization* 1(March): 39-60.

[45] Hanemann, W. M. (1991) “Willingness to pay and willingness to accept: how much can they differ?”. *American Economic Review* 81: 635-647.

[46] Shogren, J. F., Shin, S. Y., Hayes, D. J., Kliebenstein, J. B. (1994) “Resolving Differences in Willingness to Pay and Willingness To Accept”. *American Economic Review* 84: 255-270.

[47] Shogren, J. F., Cho, S., Koo, C., List, J., Park, C., Polo, P., Wilhelmi, R. (2001) “Auction Mechanisms and the Measurement of WTP and WTA”. *Resource and Energy Economics* 23: 97-109.

[48] Plott, C. R., Zeiler, K. (2005) “The Willingness to Pay – Willingness to Accept Gap, The “Endowment Effect,” Subject Misconceptions and Experimental Procedures for Eliciting Valuations” *American Economic Review* 95: 530-545.

[49] Bohm, P., Lind, H. (1993) “Preference reversal, real-world lotteries, and lottery-interested subjects” *Journal of Economic Behavior and Organization* 22:327-348.

[50] Harrison, G. W. (1989) “Theory and Misbehaviour of First-Price Auctions”. *American Economic Review* 79(September): 749-763.

[51] Lusk, J. L., Alexander, C., Rousu, M. C. (2007) “Designing Experimental Auctions for Marketing Research: The Effect of Values, Distributions, and Mechanisms on Incentives for Truthful Bidding”. *Review of Marketing Science* 5(1).

[52] Kaas, K. P., Ruprecht, H. (2006) “Are the Vickrey Auction and the BDM Mechanism Really Incentive Compatible? Empirical Results and Optimal Bidding Strategies in Cases of Uncertain Willingness-to-pay”. *Schmalenback Business Review* 58: 37-55.

[53] Wertenbroch, K., Skiera, B. (2002) “Measuring Consumers’ Willingness to Pay at the Point of Purchase”. *Journal of Marketing Research* 39: 228-241.

[54] Horowitz, J. K. (2006) “The Becker-DeGroot-Marschak mechanism is not necessarily incentive compatible, even for non-random goods”. *Economics Letters* 93: 6-11.

[55] Machina, M. (1982) “Expected utility analysis without the independence axiom”. *Econometrica* 50: 277-323.

[56] Fiorillo, C. D., Newsome, W. T., Schultz, W. (2008) “The temporal precision of reward prediction in dopamine neurons”. *Nature Neuroscience* 11: 966-973.

[57] Guitart-Masip, M., Beierholm, U. R., Dolan, R., Duzel, E., Dayan, P. (2011) “Vigor in the face of fluctuating rates of reward: an experimental examination”. *Journal of Cognitive Neuroscience* 23:3933-3938.

[58] Dickinson, A., Balleine, B. (1994) “Motivational control of goal-directed action”. Animal Learning and Behavior 22(1):1-18.

[59] Stoet, G., Snyder, L. H. (2003) “Executive control and task-switching in monkeys”. *Neuropsychologia* 41: 1357-1364.

[60] Rogers, R. D., Monsell, S. (1995) “Cost of a predictable switch between simple cognitive tasks”. *Journal of Experimental Psychology: Human Perception and Performance* 124: 207-231.

[61] Box, G. E. P., Draper, N. R. (1987) “Empirical Model-Building and Response Surfaces”. *Wiley, New York* p.424.

[62] Pearce, J. M., Hall, G. (1980) “A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli”. *Psychological Review* 87: 532-552.

[63] Esber, G. R., Haselgrove, M. (2011) “Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning”. *Proceedings of the Royal Society of Biological Sciences* 278: 2553-2561.

[64] Holland, P. C., Schiffino, F. L. (2016) “Mini-review: prediction errors, attention and associative learning”. *Neurobiology of Learning and Memory* 131: 207-215.

[65] Bouton, M. E., Sunsay, C. (2003) “Importance of trials versus accumulating time across trials in partially reinforced appetitive conditioning”. *Journal of experimental psychology: Animal Behaviour Processes* 29(1): 62-77.

[66] Harris, J. A., Patterson, A. E., Gharaei, S. (2015) “Pavlovian conditioning and cumulative reinforcement rate”. *Journal of experimental psychology: Animal learning and Cognition* 41: 137-151.

[67] Mazur, J. E. (1987) “An adjusting procedure for studying delayed reinforcement”. In: Commons, M.L., Mazur, J.E., Nevin, J.A., Rachlin, H. (Eds.), Quantitative analysis of behavior: The effects of delay and intervening events on reinforcement value, vol. 5. *Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey* p.55-73.

[68] Kobayashi, S., Schultz, W. (2008) “Influence of reward delays on responses of dopamine neurons”. *The Journal of Neuroscience* 28(31): 7837-7846.

[69] Watanabe, M., Cromwell, H. C., Tremblay, L., Hollerman, J. R., Hikosaka, K., Schultz, W. (2001) “Behavioral reactions reflecting differential reward expectations in monkeys”. *Experimental Brain Research* 140: 511-518.

[70] Enomoto, K., Matsumoto, N., Nakai, S., Satoh, T., Sato, T. K., Ueda, Y., Inokawa, H., Haruno, M., Kimura, M. (2011) “Dopamine neurons learn to encode the long-term value of multiple future rewards”. *PNAS USA* 108: 15462-15467.

[71] Crespi, L. P. (1942) “Quantitative variation of incentive and performance in the white rat”. *American Journal of Psychology* 55: 467-517.

[72] Bower, G. H., Miller, N. E. (1980) “Effects of amount of reward on strength of approach in an approach-avoidance conflict”. *Journal of comparative and physiological psychology* 53: 59-62.

[73] Rolls, B. J., Rolls, E. T. (1982) “Thirst”. *Cambridge University Press* Cambridge, New York, NY.

[74] Saker, P., Farrell, M. J., Adib, F. R. M., Egan, G. F., McKinley, M. J., Denton, D. A. (2014) “Regional brain responses associated with drinking water during thirst and after its satiation”. *PNAS USA* 111: 5379-5384.

[75] Stellar, J. R., Stellar, E. (1985) “The Neurobiology of Motivation and Reward”. *Springer-Verlay* New York.

[76] Bouret, S. Richmond, B. J. (2010) “Ventromedial and orbitofrontal prefrontal neurons differentially encode internally and externally driven motivational values in monkeys” *The Journal of Neuroscience* 30(25): 8591-8601.

[77] Aitta-aho, T., Phillips, B. U., Pappa, E., Hay, Y. A., Harnischfeger, F., Heath, C. J., Saksida, L. M., Bussey, T. J., Apergis-Schoute, J. (2017) “Accumbal cholinergic interneurons differentially influence motivation related to satiety signalling”. *eNeuro* 4(2): e0328-16.

[78] Heitz, R. P., Schall, J. D. (2012) “Neural mechanisms of speed-accuracy tradeoff” Neuron 76(3): 616-628.

[79] Pastor-Bernier, A., Plott, C. R., Schultz, W. (2017) “Monkeys choose as if maximizing utility compatible with basic principles of revealed preference theory”. *PNAS USA* 1114(10): E1766-E1775.

[80] Spence, K. W., Norris, E. B. (1950) “Eyelid conditioning as a function of the inter-trail interval”. *Journal of Experimental Psychology* 40(6): 716-720.

[81] Gallistel, C. R., Gibbon, J. (2000) “Time, Rate, and Conditioning” *Psychological Review* 107(2): 289-344.

[82] Hernadi, I., Grabenhorst, F., Schultz, W. (2015) “Planning activity for internally generated reward goals in monkey amygdala neurons”. *Nature Neuroscience* 18: 461-469.

[83] Zangemeister, L., Grabenhorst, F., Schultz, W. (2016) “Neural basis for economic saving strategies in human amygdala prefrontal reward circuits”. *Current Biology* 26(22): 3004-3013.

[84] Samejima, K., Ueda, Y., Doya, K., Kimura, M. (2005) “Representation of action-specific reward values in the striatum”. *Science* 310: 1337-1340.

[85] Paton, J. J., Belova, M. A., Morrison, S. E., Salzman, C. D. (2006) “The primate amygdala represents the positive and negative value of visual stimuli during learning”. *Nature* 439: 865-870.

[86] Ljunberg, T., Apicella, P., Schultz, W. (1992) “Responses of monkey dopamine neurons during learning of behavioural reactions*”. Journal of Neurophysiology* 67: 145-163.

[87] Schultz, W., Dayan, P., Montague, P. R. (1997) “A neural substrate of prediction and reward”. *Science* 275: 1593-1500.

[88] Bayer, H. M., Glimcher, P. W. (2005) “Midbrain dopamine neurons encode a quantitative reward prediction error signal”. *Neuron* 47: 129-141.

[89] Sescousse, G., Redoute, J., Dreher, J. (2010) “The architecture of reward value coding in the human orbitofrontal cortex”. *Journal of Neuroscience* 30(39): 13095-13104.

[90] Elliot, R., Newman, J. L., Longe, O. A., Deakin, J. F. (2003) “Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study”. *Journal of Neuroscience* 23: 303-307.

[91] Rolls, E. T. (2000) “The orbitofrontal cortex and reward”. *Cerebral Cortex* 10(3): 284-294.

[92] Matsumoto, M., Hikosaka, O. (2007) “Lateral Habenula as a source of negative reward signals in dopamine neurons”. *Nature* 447: 1111-1115.

[93] Jhou, T. C., Fields, H. L., Baxter, M. G., Saper, C. B., Holland, P. C. (2009) “The rostromedial tegmental nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses”. *Neuron* 61: 786-800.

[94] Matsumoto, M., Hikosaka, O. (2009) “Representation of negative motivational value in the primate lateral habenula”. *Nature Neuroscience* 12: 77-84.

[95] Matsumoto, M., Hikosaka, O. (2011) “Electrical stimulation of the primate lateral habenula suppresses saccadic eye movement through a learning mechanism”. *PLoS ONE* 6(10): e26701.

[96] Milgrom, P. R., Weber, R. J. (1982) “A theory of Auctions and Competitive Bidding” *Econometrica* 50: 1089-1122.

[97] Lusk, J. L., Shogren, J. F. (2007) “Experimental auctions: Methods and applications in economic and marketing research”. *Cambridge University Press* p. 73.

[98] Brainard, D. H. (1997) “The psychophysics toolbox”. *Spatial Vision* 10: 433-436.

[99] MATLAB and Statistics Toolbox Release 2012b, *The MathWorks, Inc., Natick, Massachusetts, United States*.

[100] Cox, DR (1958) "The regression analysis of binary sequences (with discussion)". *Journal of the Royal Statistical Society Series B* 20: 215–242.

[101] Zimmerman, D. W. (2004) "A note on preliminary tests of equality of variances". *British Journal of Mathematical and Statistical Psychology* 57: 173–181.