Paradoxical Choice and the Value of Information in Animals



Victor Ajuwon

St Anne's College & the Department of Biology

University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Hilary Term 2023

Acknowledgements

I've had the incredible fortune to work with amazing and kind people throughout my DPhil, who have become dear friends and mentors. My huge thanks to Andres, the rat whisperer, who brought me round to nearly enjoying working with these incredibly intelligent animals. It was such a joy to work and think with you. My hearty thanks to Tiago, my fish whispering comrade. The pandemic wasn't fun but managing to teach goldfish to race around a tank was quite a highlight. I'm looking forward to more adventures. Thanks also to Adelaide for kindly sharing your fishy expertise. I'm incredibly grateful to my supervisors, Alex, Mark and Graham, for their mentorship and insight throughout. Though I've been fascinated by animal behaviour for a long time, Alex's undergraduate lectures showing that behaviour could be predicted with almost unbelievable accuracy using Darwinian logic amazed me and are no doubt a reason that I've followed this line of research. Thank you, Alex, for your generosity of time, great humour and kindness. I've been incredibly lucky. I would also like to thank Tim Guilford, my undergraduate tutor, whose inspiring teaching helped me onto this path of teaching and research. Thanks also to Martin and his Lab in Toulouse for hosting me and teaching me the way of the bee. It was an amazing experience, and I couldn't have been more welcomed.

I would also like to thank my funders, the Biotechnology and Biological Sciences Research Council (BBSRC), who gave me the opportunity to pursue doctoral research with all the required resources. The Doctoral Training Programme was an excellent way to meet like-minded researchers in a supportive environment. My thanks also to the Department of Biology, Merton College, and St Anne's College, whose facilities and financial support were very helpful, the latter via a Graduate Development Scholarship. I am also very grateful to Columbia University's Zuckerman Institute for a generous Travel Award that enabled me to attend a memorable conference there.

My time doing research has been enriched by a range of other academic-related activities in Oxford. Atreyi, Lauren, Sara, setting up the BIPOC STEM Network was such a great achievement, and made all the better doing it with you. Thanks to Misha and Emily for carrying the torch forward! Mark and Sara, organisng the BEE conference was an epic highlight, thanks for being great to work with. And Hope, I'm so proud of Thinking Black – thanks for sharing your vision with me.

More friends than I can list here have inspired and supported me throughout, and I'm so grateful. Tom, Atreyi and Pippa, nothing beats the catharsis of those pub chats. Mark, and Tom, time spent together in Cowley was short, but not half bad. Tom, thanks for making lockdown endurable. And thanks Maisie and JT for putting a smile on my face whenever I see you. Joe, Thomas and Clem, what an

absolute hoot it’s been - thanks for everything. Joe, not too long ago we were dreaming of this! Best not to count the number of pints in between, but I’m sure every one of them was class. Thanks to Ollie, the 59ers and Rose Laners who have always been so encouraging. Thanks, Simon for your kindness and example, to Leah, Jack, Mary and Ant for the laughs, and making me feel so young. And of course, to Tristan – it’s been epic.

Thank you to my family, for your support and encouragement always, and for seeing so much in me. And finally, thanks to Jack, for all your love and support; for sharing the best times with me.

Thesis Abstract

Normative accounts of decision-making posit that information derives value by enabling decision-makers to efficiently acquire beneficial external commodities such as food. However, an experimental protocol referred to as 'paradoxical choice' shows that a range of bird and mammal species, including humans, appear to value information about upcoming events that they cannot use, as if their information-seeking behaviour was driven by an intrinsic drive to resolve uncertainty, or 'curiosity'. Using rats and goldfish as models, I investigate the paradoxical choice phenomenon experimentally and interpret it theoretically.

In Chapter 1, I use three variations of the paradoxical choice protocol in rats to investigate the psychological mechanisms underlying information-seeking behaviour. My results indicate that an intrinsic drive to resolve uncertainty, as well as a drive to generate signals for appetitive outcomes, may act simultaneously to generate preferences.

In Chapter 2, I report three experiments designed to examine how preferences for advanced information about upcoming rewards influence decision-making under risk in rats. I found that while subjects are generally risk averse, the provision of informative stimuli in the risky alternative promotes risk proneness.

In Chapter 3, I present *GoFish*, a novel, open-source platform, that facilitates automated and closed-loop learning experiments on fish and other aquatic species. As an illustration and test of its use, I present the results of 2 experiments on discrimination learning, reversal, and choice in goldfish.

In Chapter 4, I use *GoFish* and the paradoxical choice protocol to explore non-instrumental information-seeking in fish (goldfish) for the first time. Like in birds and mammals, fish learned the programmed contingencies and showed rational preferences between stimuli predicting differing reward probabilities, but, unlike the other vertebrates studied so far, this did not translate into a preference for signalled outcomes.

Together, my results expose the extent to which purely reward-maximizing theorizing is insufficient to account for behaviour. This thesis highlights the need to interlink functional and mechanistic accounts of value-based decision-making incorporating empirical insights about information-seeking, in order to understand how natural selection has shaped behaviour.

“The fact is, that the public have an insatiable curiosity to know everything, except what is worth knowing.”

-Oscar Wilde

Contents

THESIS ABSTRACT **1**

CONTENTS **4**

GENERAL INTRODUCTION **7**

PARADOXICAL CHOICE: ORIGINS AND IMPLICATIONS 12
THESIS AIMS AND STRUCTURE 21

CHAPTER 1. PARADOXICAL CHOICE AND THE REINFORCING VALUE OF INFORMATION **24**

ABSTRACT 25
INTRODUCTION 27
METHODS 39
Subjects *39*
Apparatus *39*
Training *40*
Experimental procedures *43*
Data Analysis *47*
Ethical note *48*
RESULTS 49
Training *49*
Experiment *51*
DISCUSSION 59
SUPPLEMENTARY MATERIALS 73

CHAPTER 2. NON-INSTRUMENTAL INFORMATION PROMOTES RISK-PRONENESS FOR AMOUNTS OF FOOD IN RATS **75**

ABSTRACT 76
INTRODUCTION 78
Risk Sensitivity Theory *80*
Non-instrumental information-seeking *83*
Risk and Information *86*
METHODS 89
Subjects *89*
Apparatus *89*
Cues *90*
Training *91*
Experimental procedures *93*
Data Analysis *96*
Ethical Note *97*
RESULTS 99
Cue Training *99*
Experiment *102*
DISCUSSION 105

CHAPTER 3. *GOFISH*: A LOW-COST, OPEN-SOURCE PLATFORM FOR CLOSED-LOOP BEHAVIOURAL EXPERIMENTS ON FISH **116**

ABSTRACT 117
INTRODUCTION 119
METHODS 124
The GoFish platform *124*
Pellet dispensers *129*
Stimuli *130*
Behavioural task control *131*

Fish tracking *132*
Subjects *134*
Pre-training *135*
Experiment 1: Acquisition and reversal of a spatial conditioning *137*
Experiment 2: Colour discrimination *138*
Data analysis *140*
Ethics statement *140*
RESULTS AND DISCUSSION 142
GENERAL DISCUSSION 148

CHAPTER 4. EXPLORING NON-INSTRUMENTAL INFORMATION-SEEKING IN GOLDFISH 154

ABSTRACT 155
INTRODUCTION 157
METHODS 163
Subjects *163*
Apparatus and task control *164*
Stimuli *167*
General Procedure *168*
Pre-training *169*
Experiment 1: Paradoxical choice *169*
Experiment 2: Stimulus choice *171*
Fish Tracking *172*
Data analysis *173*
Ethics statement *174*
RESULTS 176
Goldfish spatial location shows discrimination of stimulus contingencies *176*
Goldfish choice is modulated by stimulus outcomes *179*
Goldfish show no preference for advanced information in paradoxical choice *182*
DISCUSSION 185

GENERAL DISCUSSION 197

REFERENCES 217

General Introduction

Uncertainty is a ubiquitous challenge for living organisms. Across contexts diverse as foraging, navigation and mate selection, uncertainty impairs the ability of animals to effectively exploit the potential opportunities available to them and exposes individuals to risks that can have catastrophic consequences for survival and reproduction. To deal with this challenge, animals have a range of strategies to acquire biologically relevant information (Gottlieb et al., 2013). We might consider that a very simple type of information-seeking strategy involves an innate behavioural action sequence aimed at sampling the environment to encounter novel stimuli. For example, when placed in a new foraging patch, the nematode worm *C. elegans* will explore locally for a short period of time before abruptly switching to explore a much larger area, a strategy consistent with the maximization of information about the distribution of food within the environment (Calhoun et al., 2014). In laboratory choice tasks, animals also show the capacity to learn to acquire information in order to improve reward outcomes, showing that they also have the ability to flexibly adapt their behaviour in order to acquire information (Foley et al., 2017; Gottlieb et al., 2014; Kobayashi and Hsu, 2019).

The fact that individuals can learn to acquire information in order to improve reward gains is well accounted for by normative theories of value based decision-making spanning fields such as psychology (Caplin and Leahy, 2001; Rachlin et al., 1981), economics (Gravelle and Rees, 2004; Mas-Colell et al., 1995), and behavioural ecology (Stephens and Krebs, 1986). These theories postulate that individuals are rational agents aimed at consistently maximising beneficial external commodities, or utility. In evolutionarily inspired models of animal behaviour, the ultimate commodity that individuals are assumed to maximise is their reproductive success, or inclusive fitness, due to the action of natural selection (Grafen, 2009; Parker and Smith, 1990). Within the context of foraging, it is broadly contended that to achieve this individuals should maximise their average net-rate of energy intake, because all else being equal, this will lead to a greater amount of energy and time available for reproduction (see Kacelnik, 2006; Stephens and Krebs, 1986). From this perspective, information derives value from its instrumental potential i.e., the fact that it can be used by decision-makers to enhance rate of energy intake and thereby fitness (Dall et al., 2005). This view is complemented by classical models of reinforcement learning (Rescorla and Wagner, 1972) that provide a mechanistic account of how such preferences can be acquired: agents should increase the frequency of actions that ultimately result in a higher probability of essential biological commodities like food, including actions resulting in useful information about food.

Over the last few decades however, in a paradigm referred to as 'paradoxical' or 'sub-optimal' choice, a growing range of mammal and bird species, including humans, have been shown to learn to choose alternatives that provide non-instrumental information – information that cannot improve reward gains in the laboratory– even though doing so can lead to significant losses – in some cases up to 80% of potential available rewards (for reviews see Cunningham and Shahan, 2018; Dunn et al., 2023; Zentall, 2016). While traditional, theories of decision-making can explain the fact that individuals may forgo food in the short-term to seek out information that will enhance their long-term gains (a trade-off labelled the exploration-exploitation dilemma (Berger-Tal et al., 2014; Cohen et al., 2007; Krebs et al., 1978)), these accounts appear challenged by the paradoxical finding that individuals are motivated to acquire costly information that cannot be used to adaptively modify behaviour. Subsequently, a rich and active debate concerning the theoretical interpretation of these intriguing preferences has developed, spanning neuroscience (Bromberg-Martin and Monosov, 2020; Gottlieb and Oudeyer, 2018), behavioural ecology (Vasconcelos et al., 2018, 2015), economics (Loewenstein and Molnar, 2018) and experimental (Dunn et al., 2023) and cognitive (Kidd and Hayden, 2015) psychology.

In this thesis, my primary goal is to investigate how and why in the laboratory animals learn to seek advanced information that has no instrumental benefit - as if information were valued itself - using variations of the paradoxical choice task. My specific aims are to (i) further our understanding of the psychological mechanisms underlying preferences for advanced non-instrumental information and consider their adaptive implications, (ii) more broadly investigate the interaction of non-instrumental information-seeking with other aspects of decision-making, with a particular focus on risk, and (iii) comparatively examine the ubiquity of preferences for non-instrumental information within the vertebrate clade. First, to provide background context, I will map the origins of the paradoxical choice paradigm and survey the theoretical and experimental developments that have followed, before outlining the aims of the chapters of this thesis.

Paradoxical choice: origins and implications

Almost 75 years ago, Wyckoff (1951 unpublished thesis; see Wyckoff 1969) presented pigeons with a task in which subjects could produce a response that would generate stimuli informing them about the availability of food. He labelled this response - the pressing of a pedal - the 'observing response' because it allowed subjects to observe reward predictive stimuli across trials in where possibility of forthcoming rewards was uncertain. In randomly intermixed trials the pecking of a key resulted in a 50% chance of reward after a 30 s period. By pressing on the pedal, subjects could inform themselves of the forthcoming outcome as this would result in one colour if food would be available, or another if not. Therefore, the observing response reduced reward related uncertainty, but did not alter the amount of food available to the pigeons. Wyckoff found that his subjects readily acquired the observing response. Though he was initially interested in how pigeons might learn to discriminate between the two signals generated by the response, the pertinent question soon became why the pigeons sought to observe the reward predictive stimuli in the first place, given that they could not use the information to increase reward rate.

To explore the mechanistic basis of observing responses a number of research groups examined the relative contribution that the signal for forthcoming reward, and the signal for no reward have in the acquisition and maintenance of observing responses (reviewed in Dinsmoor, 1983). The general finding was that observing responses were acquired when they exclusively generated explicit signals for forthcoming reward, but not when they only generated signals for no reward. These findings had important implications for theoretical accounts of observing responses, which are discussed in detail in the first Chapter.

Following Wyckoff's finding, Prokasy (1956) carried out a similar procedure in rats, but in this case, subjects had to explicitly make a choice between one of two alternatives in order to observe reward predictive signals. Prokasy used a maze in the shape of an 'E', where subjects started in a central chamber and could make a choice by turning either left or right to move into one of the two choice arms. After a choice, subjects waited in a delay chamber for 30 s before being allowed to enter a goal box where there was a 50% chance of reward. One of the choice arms was informative, in that the colour of the delay chamber (black or white) was correlated with the forthcoming reward outcome (food or no food), while in the other choice arm the colour of the delay chamber was not correlated with outcomes. Like Wyckoff, Prokasy found that most subjects engaged in observing behaviour by choosing the informative option.

The paradoxical choice protocol (also referred to as sub-optimal choice), which is employed in this thesis, is in essence very similar to, and derivative of observing response protocols like Prokasy's. In the paradoxical choice task, which was originally developed in pigeons (for reviews see Cunningham and Shahan, 2018; Dunn et al., 2023; Zentall, 2016) individuals must choose between one of two manipulanda (e.g., pecking key or lever), which both result in a short delay (typically 10s) and then a probability of receiving one of two reward outcomes, usually either reward or no reward. One of the options is informative, producing reward-predictive signals immediately after the choice response, while the other alternative does not provide informative signals, so subjects are uncertain about reward outcomes until the delay has lapsed. A distinguishing feature of the paradoxical choice task is that subjects often have to sacrifice substantial rewards for advanced information, because the informative alternative provides less average reward than the non-informative one (though in this thesis alternatives are equally profitable). The critical finding is that despite potentially incurring reward losses, pigeons, starlings, rats, and monkeys nonetheless show robust preferences for the informative alternative, suggesting that subjects attribute subjective value to the acquisition of non-instrumental information.

At the time that the observing response was first documented, two theoretical explanations were put forward (reviewed in Mackintosh, 1974) that have endured as putative underlying mechanisms. More recently these theories have been

elaborated and expanded on to account for non-instrumental information-seeking behaviour across a range of species, including humans. Both theories assume that agents ascribe subjective value to reward-predictive stimuli. One proposal is that information, or a reduction in uncertainty about upcoming rewards, is it itself intrinsically valuable, in the same way that food rewards are (Berlyne, 1960; Hendry, 1969). Hence subjects perform the observing response to generate informative stimuli. The implication is that information can act as a primary reinforcer in learning processes. Information is a complex concept with a range of definitions. At the time it was assumed that information was quantified by subjects in a manner consistent with Shannon's Information Theory (1948), where information is defined by its ability to reduce the number of possible outcomes associated with an uncertain events (see MacKay, 2003) e.g., the potential delivery of food rewards.

There are a number of ways in which information may act to reinforce preferences that generate reward-predictive stimuli. One possibility is that information is reinforcing because it alleviates temporally prolonged uncertainty, which is aversive (Bennett et al. 2016). From this view, information modulates preference as a negative reinforcer: by eliminating or reducing the presence of an aversive state, reward-predictive stimuli conveying information reinforce preferences. On the other hand, information, or the resolution of uncertainty, may act as a positive reinforcer, where information is itself appetitive, a notion consistent with active

inference theory (Friston et al. 2013) and dopaminergic bonuses elicited by reward predictive stimuli (e.g., Bromberg-Martin and Hikosaka, 2009). According to Shannon's information theory (1948), uncertainty reduction and information gain are mathematically equivalent, nonetheless the reinforcing nature of information has important behavioural consequences. For example, while a positively reinforcing effect of information gain would likely encourage exploration in natural settings, uncertainty aversion may have the opposite effect.

The alternative account is based on conditioning of reward predictive stimuli, which are assumed to acquire the ability to modulate responses via learned association with reward outcomes (Prokasy, 1956; Wyckoff, 1959). According to this theory, stimuli announcing 'good news' or positive outcomes reinforce responses to a greater extent than 'bad news' stimuli inhibit responses. Thus, subjects perform observing responses and select alternatives with reward predictive cues not to reduce uncertainty *per se,* but rather to generate good news stimuli that acquire appetitive properties. Results from observing response studies on pigeons designed to disambiguate these theories were interpreted to suggest that the conditioning account was more likely than a primary reinforcing effect of uncertainty reduction (Dinsmoor, 1983; Roper and Zentall, 1999; Shahan and Cunningham, 2015).

Subsequently, within the experimental psychology literature, efforts have mostly been directed at developing quantitative, models aimed at explaining how signals for good news can acquire reinforcing properties that can lead subjects to perform observing responses and develop sub-maximal preferences in the paradoxical choice task (reviewed in Cunningham and Shahan, 2018; Dunn et al., 2023; Zentall, 2016; see also Beierholm and Dayan, 2010; González et al., 2020).

Recently, the observing response and paradoxical choice paradigms have inspired a number of theoretical and empirical studies across a range of fields and species that have led to a resurgence in research aimed at understanding the functional and mechanistic basis of non-instrumental information-seeking behaviour. These increasingly interdisciplinary efforts have enhanced the scientific value and impact of the protocol and helped to develop new theoretical frameworks with the promise of further advancing our understanding of how natural selection has shaped decision-making mechanisms.

Several observing response-inspired studies using neuroscientific and computational approaches have reinvigorated the decades-old debate regarding whether information can act as a primary reinforcer (Bromberg-Martin and Monosov, 2020), while also suggesting that preferences for non-instrumental information may reflect a rudimentary form of curiosity (Cervera et al., 2020; Gottlieb and Oudeyer, 2018; Kidd and Hayden, 2015; Kobayashi et al., 2019).

One of the key findings is that neurones in the brains of macaques responsible for representing the value of primary rewards such as food and water also produce similar signals for advanced information about upcoming food rewards (Bromberg-Martin and Hikosaka, 2011, 2009), suggesting that at least in macaques, information may act as a primary reinforcer in learning processes. This finding is consistent with paradoxical choice- inspired experiments on human subjects using computational approaches showing that information may derive value by allowing subjects to avoid uncertainty which is aversive (Bennett et al., 2016; Brydevall et al., 2018). However, recent observing response-inspired experiments by Iigaya et al. (2020, 2016) have used reinforcement learning approaches to show that preferences for informative alternatives may instead be driven by the savouring of positive future outcomes boosted by signals for good news. This idea is informed by the concept of anticipatory utility in economics (Kreps and Porteus, 1978), and is similar to earlier proposals emphasising the importance of signals for good news acquiring appetitive properties through conditioning. Like some other authors (Zentall et al., 2019; Zentall and Stagner, 2011), Ligaya et al. (2016) argue that preferences in paradoxical choice experiments may reflect mechanisms implicated in human gambling rather than the valuing of uncertainty reduction. Taken together, these results highlight that further studies are needed to fully elucidate the reinforcing properties of advanced information and its broader behavioural implications.

Paradoxical choice studies on a broader range of species have also helped to further our understanding of the functional significance of non-instrumental information-seeking. In particular, studies on starlings extending Optimal Foraging Theory (Stephens and Krebs, 1986) indicate that preferences for advanced information leading to submaximal outcomes in the lab, would actually maximise reward rates in natural situations, because information pertaining to rewards will most often be useable by subjects, unlike in the artificial experimental protocol (Vasconcelos et al., 2018, 2015). Similar arguments have been made suggesting that rather than being 'sub-optimal' as some have claimed (Zentall et al., 2019; Zentall and Stagner, 2011),

preferences for advanced non-instrumental information reflect the fact that in natural settings open-ended information-seeking would help agents build an accurate model of the world (Bennett et al., 2016; Bromberg-Martin and Hikosaka, 2009; Bromberg-Martin and Monosov, 2020). The discovery of similar, but somewhat weaker preferences in rats (Ojeda et al., 2018; see also Ojeda 2018 unpublished thesis) is consistent with these views: the prevalence of paradoxical choice in both mammals and birds - two distantly related taxa - implies that the mechanisms underlying these preferences are adaptive in natural settings. These comparative studies have also helped to highlight interspecies differences in information-seeking behaviour, while raising questions about the evolutionary origins and ubiquity of information value across animals.

Though the paradoxical choice paradigm's origins can be traced back to almost 75 years ago, it continues to raise fundamental, interdisciplinary questions about the mechanisms underlying information-seeking behaviour, how and why they originated, and the extent to which information value more broadly influences the behaviour of decision-makers. In this 'disinformation age' (Livingston and Bennett, 2020) tackling these questions is more important than ever, and will help us better understand how natural selection has shaped decision-making mechanisms, with implications for our understanding of human health and cognition, and potentially the development of artificial general intelligence (Kaplan and Oudeyer, 2007; Oudeyer et al., 2007).

Thesis Aims and Structure

This thesis aims to further our understanding of non-instrumental information-seeking behaviour by investigating the mechanisms that underly it, how information-seeking behaviour influences decision-making, and how ubiquitous non-instrumental information-seeking is across distantly related taxa, using the paradoxical choice task. The thesis is organised into four manuscript chapters, the first and third of which are published articles. In the first two chapters I investigate the mechanisms underlying non-instrumental information-seeking behaviour in rats and explore its effects on decision-making more broadly by examining how information-seeking influences preferences for risk. In the last two chapters I explore the paradoxical choice phenomenon in fish for the first time. I present a new platform for automated learning experiments on aquatic species, demonstrating its potential and utility with goldfish. Then, using this platform I investigate the prevalence of non-instrumental information-seeking in goldfish.

In Chapter 1, I aim to disambiguate between two putative psychological mechanisms that may drive non-instrumental information-seeking in paradoxical choice, using rats as model subjects. The information hypothesis proposes that individuals seek to resolve uncertainty, while the conditioned reinforcement

hypothesis argues that preferences are motivated by a drive to generate appetitive 'good news' stimuli. Across three groups I test a distinguishing prediction between the hypotheses concerning the ability of stimuli announcing 'bad news' to reinforce information-seeking behaviour.

In Chapter 2, I examine the influence that non-instrumental information-seeking has on decision-making more broadly by examining how this behaviour influences decision under risk in rats. Risk is particularly interesting in this context because generally animals are found to avoid alternatives that are risky in reward amount, but in the paradoxical choice literature, under experimental conditions subjects have been shown to prefer an informative risky alternative over a fixed uninformative alternative one. To address how individuals trade-off these determinants of preference, across three experiments I investigate the independent effects that advanced information and risk have on preferences, and then examine how both of these effects interact to generate preference.

With the ultimate goal of exploring non-instrumental information-seeking in fish using refined procedures, in Chapter 3 I develop a platform that can facilitate fully automated, closed-loop learning experiments on fish, demonstrating its utility using classical discrimination learning paradigms.

In Chapter 4, I explore the ubiquity of information value in vertebrates, by implementing the paradoxical choice protocol using a fish (goldfish), to my knowledge, for the first time.

Chapter 1. Paradoxical choice and the reinforcing value of information

A manuscript published in Animal Cognition

Abstract

Signals that reduce uncertainty can be valuable because well informed decision-makers can better align their preferences to opportunities. However, some birds and mammals display an appetite for informative signals that cannot be used to increase returns. We explore the role that reward-predictive stimuli have in fostering such preferences, aiming at distinguishing between two putative underlying mechanisms. The 'information hypothesis', proposes that reducing uncertainty is reinforcing *per se,* somewhat consistently with the concept of curiosity: a motivation to know in the absence of tractable extrinsic benefits. In contrast, the 'conditioned reinforcement hypothesis', an associative account, proposes asymmetries in secondarily acquired reinforcement: post-choice stimuli announcing forthcoming rewards (S^+) reinforce responses more than stimuli signalling no rewards (S^-) inhibit responses. In three treatments, rats faced two equally profitable options delivering food probabilistically after a fixed delay. In the informative option (*Info)*, food or no food was signalled immediately after choice, whereas in the non-informative option (*NoInfo*) outcomes were uncertain until the delay lapsed. Subjects preferred *Info* when (1) both outcomes were explicitly signalled by salient auditory cues, (2) only forthcoming food delivery was explicitly signalled, and (3) only the absence of forthcoming reward was explicitly signalled. Acquisition was slower in (3), when food was not explicitly

signalled, showing that signals for positive outcomes have a greater influence on the development of preference than signals for negative ones. Our results are consistent with an elaborated conditioned reinforcement account, and with the conjecture that both uncertainty reduction and conditioned reinforcement jointly act to generate preference.

Introduction

Models of instrumental learning in animals, and of reinforcement learning in machines, argue that agents increase the frequency of actions that result in higher probability of desirable consequences, and reduce the frequency of actions with undesirable ones. Desirability and aversiveness have adaptive roots in animals (for a review see Staddon and Cerutti, 2003), while in machines reinforcement criteria are built in by design (Sutton and Barto, 2018). In animals, constrained essential commodities or other substantial beneficial outcomes are effective reinforcers, and it is not surprising that in laboratory studies signals for their occurrence or absence modulate animals' lever pressing or key-pecking responses. However, a question that has recently seen a resurgence of interest in the psychological (Cunningham and Shahan, 2018; Shahan and Cunningham, 2015), neuroscientific/robotics (Gottlieb and Oudeyer, 2018; van Lieshout et al., 2020), and computational (Dubey and Griffiths, 2020) literatures is whether in addition to such commodities, information (reductions in uncertainty) can modulate behaviour through the same processes that conventional rewards do, that is, whether information can act as a primary reinforcer.

In a world where uncertainty is pervasive, information is a valuable asset that can be used by decision-makers to enhance efficiency in activities such as foraging,

mating or homing, to improve their performance (Behrens et al., 2007; Dall et al., 2005) and ultimately contribute to Darwinian fitness. In instrumental tasks, animals may seek information before making choices (Gottlieb et al., 2014) and this can improve the acquisition of commodities (Foley et al., 2017; Kobayashi and Hsu, 2019). In such contexts, the adaptive and reinforcing values of information-seeking derive from its ability to increment a well-defined benefit, in which case traditional functional and mechanistic accounts are aligned.

However what seems paradoxical with reference to normative models of reward-maximisation across fields such as microeconomics and foraging theory (e.g., Mas-Colell et al., 1995; Stephens and Krebs, 1986), and classical models of reinforcement learning (Rescorla and Wagner, 1972), is that animals show preferences for informative signals in cases where the signals cannot increase reward rate, indicating that subjects seek out information 'for its own sake', are 'uncertainty averse', or are 'curious' (Bromberg-Martin and Hikosaka 2009; Cervera et al., 2020; Kidd & Hayden, 2015). The idea that animals value information irrespective of its ability to increase rewards was postulated to explain the experimental phenomenon of the 'observing response', first explored by Wyckoff (1951, *unpublished thesis;* see Wyckoff, 1969). In this paradigm, subjects can resolve uncertainty about forthcoming contingencies by performing a response, though the information provided cannot be used to modify outcomes. Wyckoff presented pigeons with a white key and a mixture of two reinforcement

schedules. In the rich schedule, pecking the key resulted in food delivery every 30s, while in the poor schedule pecking at the key did not produce food. The system alternated periodically and unpredictably between both schedules. The critical aspect was the addition of a pedal such that if the animal stepped on it, then the white key turned red when the system was in its rich state and green during the poor periods. The pedal response informed the animal of the current state of the world but did not modify it. The pigeons readily acquired pedal pressing, which was appropriately labelled an 'observing response'. Similar procedures have been conducted with variable delays to food (Bower et al., 1966) and aversive outcomes such as electric shocks (Lockard, 1963). In all cases animals acquire such observing responses; they choose to elicit signals that resolve uncertainty about probabilistic future outcomes, without altering these outcomes.

A number of theoretical hypotheses have been proposed to explain apparent information-seeking behaviour in the observing response paradigm (for a review see Dinsmoor,1983) and other protocols derived from it. One candidate mechanistic explanation, which we will call the 'information' or 'uncertainty reduction' hypothesis - informed by classical information theory (Shannon 1948) - suggests that animals find information itself intrinsically rewarding potentially because it relieves uncertainty, which has negative hedonic valence (Berlyne, 1960, 1957; Hendry, 1969), or because the resolution of uncertainty is appetitive

(Lowenstein, 1994). According to this account, information (and by extension an event or stimulus that reduces uncertainty) acts as a primary reinforcer modulating behaviour. Functionally, this could evolve if information is often associated with substantive benefits in ecological contexts and is not too costly to acquire. This view is consistent with notions of 'curiosity' defined as a motivation to 'know' for the sake of it, or acquire information in the absence of instrumental incentives (Cervera et al., 2020; Gottlieb and Oudeyer, 2018; Kidd and Hayden, 2015). The idea that individuals value information *per se* has also recently been explored in humans. Bennet et al. (2016) suggested that information may be valued because it prevents temporally-prolonged uncertainty, which is presumed to be aversive. Other investigators have proposed that information may derive its reinforcing value by enabling subjects to appetitively 'savour' good news about positive outcomes while waiting for those outcomes (Brydevall et al., 2018; Iigaya et al., 2016).

An alternative mechanistic explanation, which we refer to as the 'conditioned reinforcement hypothesis' (Bower et al., 1966; Prokasy, 1956; Wyckoff, 1959), prescinds of attributing reinforcing properties to information *per se.* It instead proposes that the signal for 'good news' (S^+) in observing response tasks acquires secondary reinforcing properties because it is paired with food and becomes a conditioned reinforcer that then supports the acquisition of the response (i.e., S^+

acquires appetitive, or excitatory properties). By definition, a reinforcer is an event that modifies the frequency of a response when the event is contingent on that response (e.g., Gallistel and Gibbon, 2000). For example, the presentation of food is a positive primary reinforcer because when it is contingent on a lever being pressed, animals will press the lever more frequently than when the lever pressing is not paired with food. Conditioned reinforcers are initially neutral stimuli (e.g. a clicker sound) that themselves become reinforcing after having been paired with a primary reinforcer (see Mackintosh, 1974, and for applications in machine learning see Sutton and Barto, 2018). Thus, it has been argued that it is S^+, once it has been associated with food, that reinforces behaviour in the observing response task and other tasks derived from it. The difficulty with this hypothesis is that by the same reasoning, the signal that is paired with a negative outcome, or 'bad news' (S^-) might be expected to become a secondary conditioner for outcome absence and acquire the power to reduce the frequency of responding (i.e., S^- acquires inhibitory properties). If these two effects were of the same magnitude, then the preference for informative signals would not be acquired. However, if S^+ increases response frequency more than S^- decreases it (which follows from the theoretical assumption that outcomes are more effective for learning than their absence Rescorla and Wagner 1972; see also Murphy et al., 2001), or indeed if S^- is ignored altogether as several models assume (e.g., Cunningham and Shahan, 2018; McDevitt et al., 2016; Smith et al. 2016),

conditioned reinforcement offers a descriptive account of the acquisition of observing responses (and similar preferences) that is not dependent on the animal being sensitive to uncertainty or its reduction.

These two hypothetical mechanisms are not, however, mutually exclusive. One postulates that information is reinforcing *per se*, and the other that signals for food can acquire secondary reinforcing properties because of their food contingency: both mechanisms are indeed biologically plausible. The issue is whether apparently paradoxical effects are better explained due to one factor or the other.

To arbitrate between both hypotheses, researchers have carried out cue manipulation experiments in which either the good (e.g. food) or bad (e.g. no food) outcome is no longer preceded by a signal, or in other words the presentation of either S^+ or S^- is omitted (Dinsmoor, 1983; Dinsmoor et al., 1972; Silberberg and Fantino, 2010). According to the information hypothesis, observing responses, or preference for an informative option, should be acquired and maintained by either S^+ or S^-, since both resolve the uncertainty, but if conditioned reinforcement is the fundamental mechanism, they should only be acquired and/or maintained if and when the signal for good news (S^+) is present. Though the information hypothesis is simple, functionally appealing, and

intuitive, it fell broadly out of favour when evidence in pigeons began to emerge that was interpreted to be incongruent with it, but consistent with the conditioned reinforcement account. These cue manipulation experiments found that S^- alone was not sufficient to maintain observing responses(e.g., Dinsmoor et al., 1972; Jenkins and Boakes, 1973; Kendall, 1973; Silberberg and Fantino, 2010). Furthermore, Roper and Zentall (1999) failed to corroborate the information hypothesis' prediction that preference for the discriminative stimuli should be maximal when the degree of uncertainty is maximal (i.e. when positive and negative outcomes are equiprobable). These results therefore lead to the broad interpretation that information gain is not sufficient to explain observed preferences (Dinsmoor, 1983; Shahan and Cunningham, 2015).

Recently however, results from protocols similar in rationale to, and derived from, the observing response protocol have rekindled interest in the possibility that animals find information intrinsically rewarding. Experiments in monkeys have found that they prefer to receive unambiguous signals about the magnitude of upcoming water rewards, over ambiguous or delayed signals, and are willing to forfeit water rewards to do so. These preferences are correlated with activity in neurons implicated in the representation of primary rewards (Blanchard et al., 2015; Bromberg-Martin and Hikosaka, 2011, 2009), suggesting an intrinsic valuation of information. Furthermore, Daddaoua et al. (2016) showed that monkeys learn to actively search for Pavlovian cues to obtain conditioned

reinforcement, and also reduce uncertainty, though it is not yet clear how generalisable this result is to other species.

Moreover, 'paradoxical' (also called 'suboptimal') choice experiments have found that pigeons (e.g., Fortes et al., 2016; González et al., 2020; Macías et al., 2021; McDevitt et al., 2018; Smith et al., 2016 also see McDevitt et al., 2016 and Zentall, 2016 for reviews), starlings (Vasconcelos et al., 2015) and rats (Cunningham and Shahan, 2019; Ojeda et al., 2018), prefer an alternative that provides information that they cannot use, not just when the information is neutral with respect to reward maximisation, but even when the informative option provides less reward. In this paradigm, both alternatives result in probabilistic food delivery after a delay. In the informative option, signals (S^+ or S^-) anticipate the trial's forthcoming outcome immediately after a choice response, while in the non-informative option subjects are uncertain about outcomes throughout the delay. Remarkably, pigeons and starlings choose the informative option when it gives 80% less reward than the non-informative alternative (Fortes et al., 2016; Vasconcelos et al., 2015), while rats can sacrifice at least 20% of potential rewards (Cunningham and Shahan, 2019; Ojeda et al. 2018) by selecting the informative option. The fact that a range of experimental animals forfeit such significant amounts of food reward to generate apparently functionless, predictive signals is a strong reason to suspect a hypothetical primary reinforcing value of uncertainty reduction, given the prevalence of uncertainty in ecological settings

and the functionally problematic alternative requiring that positively conditioned stimuli can supersede the value of their unconditioned reward contingency. Taken together therefore, currently available results show that the old conundrum of whether a reduction in uncertainty can by itself reinforce behaviour is unresolved.

To investigate whether uncertainty reduction or conditioned reinforcement may better account for information-seeking behaviour we conducted paradoxical choice experiments on rats (*Rattus norvegicus)*, manipulating the salience of reward predictive cues across three treatments. All subjects were exposed to repeated choices between two options of equal average profitability. Each option delivered reward with 50% probability, a fixed time after being chosen. In all treatments, in the informative option (*Info*), the outcome of trials (food/no food) was signalled (or otherwise predictable) between each choice and the outcome, while in the other option (*NoInfo)* the outcome remained uncertain until it was realised (however, the signalling details differed between treatments, as explained below). Because the signalling occurred post-choice in *Info*, it could not be used to modify the probability of receiving food. To put it in conditioning language, the actions of choosing *Info* and *NoInfo* were both always followed by a 50% probability of a food outcome.

In the $S^+_S^-$ treatment, the interval between choosing *Info* and the outcome was filled in rewarded or unrewarded trials by either of two sounds, namely S^+, or S^-

cues respectively. In the *Only_S*$^{-}$ treatment, the interval was silent in trials when food was due but filled with a sound when no food was coming (i.e., there was no explicit S^{+} cue). In the *Only_S*$^{+}$ treatment, the same interval was filled with a sound signal in trials when food was forthcoming and with silence when it was not (i.e., there was no explicit S^{-} cue).

The information hypothesis predicts that both S^{+} and S^{-} individually, should be sufficient to generate observing responses, because both provide information about an otherwise uncertain outcome. The conditioned reinforcement hypothesis, on the other hand, stipulates that only S^{+} should be positively reinforcing. According to this view, although S^{-} reduces uncertainty just as much as S^{+}, its presence should reduce rather than increase the acquisition of an observing response or preference for an informative option.

Strong support for the conditioned reinforcement hypothesis would be corroborated if a preference for the informative option were recorded when the signal for no reward is omitted (*Only_S*$^{+}$ treatment) but not when the signal for sure reward is omitted (*Only_S*$^{-}$ treatment), as this would show that a salient S^{+} is both necessary and sufficient for the development of *Info* preference. On the other hand, if subjects developed an equally strong preference for the informative option regardless of whether S^{+} or S^{-} were omitted, the results would be

consistent with the predictions of the information hypothesis, because both resolve uncertainty to the same degree regardless of their valence. It is worth noting that the predictions of both of these hypotheses rely on the fact that the described manipulations to signalling properties of the stimuli cause functional changes to the nature of the stimuli as perceived by the subjects, rather than the experimenter's expectation. One way we are able to gain insight into this is by examining subjects' responses to stimuli while they are presented, which can help determine whether functional stimuli have indeed been omitted.

Manipulations of the signalling properties of choice alternatives in the paradoxical choice procedure have been performed previously in pigeons and starlings (e.g., Fortes et al., 2017; Vasconcelos et al., 2015 and see McDevitt et al., 1997 for similar tasks), but we found no reports of the relative quantitative impact of symmetrical omissions of S^+ and S^-, the most distinctive prediction of the two hypotheses. Thus, our experiment offers novel insights into the putative mechanisms underlying information-seeking behaviour.

We recorded two measures of preference, namely proportion of choices in 2-option choice trials, and response latencies in 1-option forced trials. Response latencies have proven to be a robust metric of preference in a variety of different behavioural protocols and species (viz. Kacelnik et al., 2011; Monteiro et al.,

2020; Reboreda and Kacelnik, 1991; Sasaki et al., 2018; Shapiro et al., 2008; Smith et al., 2018).

Methods

Subjects

Twenty-four male Lister Hooded rats (provider Envigo), 11 weeks old at the start of the experiment served as subjects. We used an all-male cohort to reduce inter-individual variability and temporal within-subject variability, but it is worth noticing that it is possible that exploratory (information-seeking) behaviour in females varies adaptively through the estrous cycle, and this is an important topic in itself. Animals were housed in groups of four. Throughout the experiment subjects were food deprived to a minimum of 85-90% of their expected free-feeding weight using growth curves from the provider. Initial weight: 337 ± 14, final weight: 357 ± 16 (mean ± std.) Water was provided ad-libitum in their home cages, and they were maintained on a 12-hour dark/light cycle with lights on at 6 A.M.

Apparatus

Testing was carried out in eight operant chambers (Med Associates, USA.) Each chamber contained three retractable levers: one in the back panel (centre) and two in the front panel, left and right of a central food magazine delivery tray. The tray

was equipped with an infrared beam and a sensor to record head entry. Each reward delivery consisted of four 45mg sucrose pellets (TestDiet, USA). A speaker was positioned above the food magazine in the front panel. Each chamber was also equipped with a houselight (white) and a fan, which were switched on for the duration of the session. The chambers were controlled via custom-written Med-State Notation programs running on MED-PC V (Med Associates, USA).

Training

Magazine training

To habituate the rats to the box and the delivery of food rewards, training began with a single variable interval session where food was delivered on average once a minute (VI60 free food schedule) a total of 60 times. The variable interval was sampled from a truncated Poisson distribution with a mean of 60 seconds and range of 0-120 seconds.

Lever training

Over the next three sessions the rats were trained to press the two front levers. Either lever (left or right with equal probability) was available on each trial (60 trials per session). Once a lever extended into the chamber, a single press resulted

in its retraction and immediate reward delivery (Fixed Ratio 1 schedule). One of the levers then again became available after a delay composed of a constant duration plus a variable one. The constant component was 5 seconds, and the variable one was sampled from a truncated Poisson distribution with a mean of 20 seconds and a range of 5 – 60 seconds. All three sessions concluded after 60 reward deliveries i.e., 30 lever presses on each side, or after 3 hours.

Cues

Within the main experiment and training there were four auditory cues, all with a duration of 10 seconds, and each associated to a reward probability. There were two cues for the informative option: S^+ (100% reward probability) and S^- (0% reward probability), and two cues for the non-informative option: N1 and N2 (both with 50% reward probability). The four sounds were: a low frequency pure tone (3 kHz, 78 dB), a high frequency pure tone (6 kHz, 78 dB), a buzzing sound (78 dB) and a clicking sound (74 dB). Assignment of sounds to reward probabilities was counterbalanced across subjects to avoid the possibility of option preferences being influenced by any intrinsic aversive or attractive properties of the sounds.

Cue training

To train the subjects to the reward contingencies of the four auditory cues (S^+, S^-, N1, and N2), the main experiment was preceded by a Pavlovian protocol in which all the rats were exposed to the cues and their respective reward contingencies. In this phase, cue presentation was independent of the behaviour of the rat. These cue-training sessions consisted of 40 trials, with 10 trials for each of the 4 cues, intermixed in random order. To avoid large deviations from the expected outcome probabilities of cues N1 and N2 in each session, proportions of outcomes were fixed as one half for each cue. Trials were separated by an ITI generated by sampling from of a truncated Poisson distribution with a mean of 50 seconds (range: 10-120 seconds) + 10 seconds (to ensure a minimum ITI of 20s; range: 20s ≤ ITI ≤130s). Subjects performed one daily session of this phase for 10 days. Cumulative time spent head-poking into the food magazine was measured to establish the degree of cue discrimination.

Experimental procedures

General procedure

We used a trial-based chain procedure as displayed in Fig. 1. There were two kinds of trials: 2-option choice trials and 1-option forced trials. A day's session was composed of 60 trials: 40 forced (half *Info* and half *NoInfo*) and 20 choice, which were randomly intermixed. All trials started with the rear lever extending. Pressing this lever resulted in its retraction, and either one (forced trials) or both (choice trials) of the front levers being presented. Pressing a front lever could initiate an acoustic cue and the retraction of that lever (forced trials) or of both levers (choice trials). The auditory cue, if present, was broadcast for a 10 second interval, after which food delivery occurred in rewarded trials without the need for a further response. Thus, each option was programmed as a discrete trial, response initiated, fixed time 10s, partial reinforcement 50% schedule. Trials were separated by an inter-trial interval (ITI) generated by sampling from a truncated Poisson distribution with a mean of 50 seconds (range: 10 – 120 seconds) and adding 10 seconds. A session finished after 60 trials or 3 hours, whichever occurred first.

Experimental Procedure

In the $S^+_S^-$ treatment, choosing *Info* resulted with equal probability in either S^+, which was paired with reward, or S^-, which was paired with no reward, thereby reliably informing the subject of the forthcoming outcome. Pressing the *NoInfo* lever, on the other hand, resulted with equal probability in either of two cues: N1 or N2, which were both associated with a 50% probability of either outcome; therefore, neither cue informed the subject of forthcoming reward.

The other two treatments, $Only_S^+$ and $Only_S^-$, differed from the $S^+_S^-$ treatment only in the signalling properties of *Info*. In $Only_S^-$ responding to *Info*, resulted with equal probability (50%) in either a 10 second silence, followed by the delayed reward (the omission of a cue associated with reward i.e., S^+), or the auditory S^- cue, which was associated with no reward. In $Only_S^+$ choosing *Info* resulted with equal probability (50%) in either the cue S^+, which was associated with reward after 10 seconds, or a 10 second silence (omission of the S^- cue) followed by no reward.

A between-subjects design was used, with eight rats in each group. Subject assignment to group was organised such that there was no correlation between group and any of the following parameters: side of the informative option; hour

of testing; cage in which the animals were housed, or cue-reward contingencies. For each group, the subjects performed one daily session for 14 days. Each rat was trained at the same time every day; one cohort of rats began the experiment at 9:00 A.M., another at 12:30 P.M., and the last at 3:30 P.M.

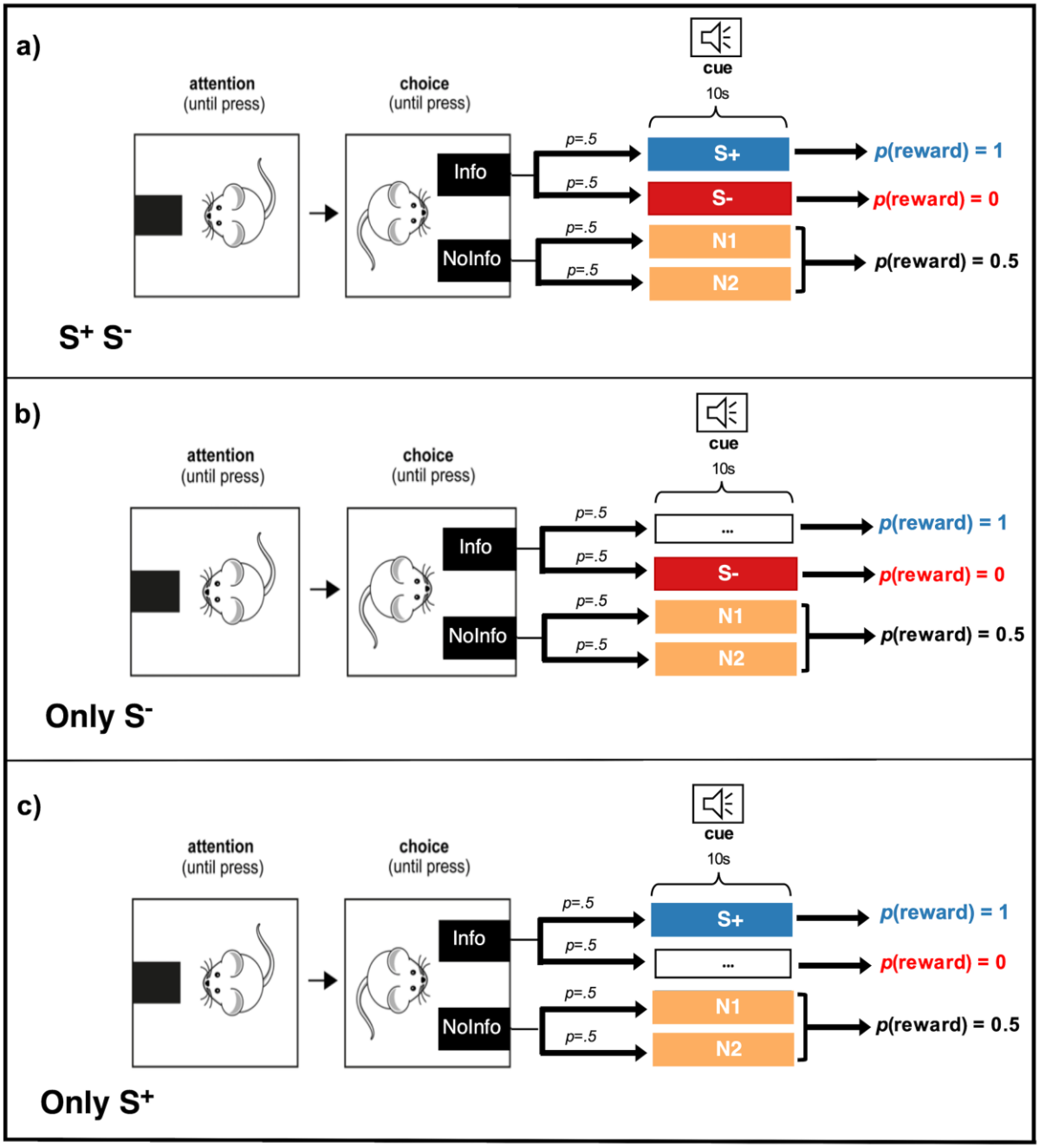


Fig. 1 Experimental design showing choice trial structure for each treatment. Blank boxes with ellipsis indicate no auditory signal (silence) preceding outcomes. *p* denotes probability.

Data Analysis

Data processing and analysis was carried out in MATLAB 2017a and statistical tests were carried out with R statistical software (https://www.r-project.org; Version 1.2.5033). A type-1 error rate of 0.05 was adopted for all statistical comparisons and the Tukey test was used for all multiple comparisons. For statistical analysis, choice proportion data were arc-sine square-root transformed to normalize the residuals. Head-poking data as well as latency index data, were square root transformed (Grafen and Hails, 2002). For all analyses of head-poking data, data from both stimuli signalling 50% chance of reward were averaged for each subject.

To measure preference on the basis of latency to respond in forced trials, for each individual we calculated an index, $L_{(Info)}$, using the median latencies to respond on *Info* ($R_{(Info)}$) and *NoInfo* ($R_{(NoInfo)}$) forced trials for each session: $L_{(Info)} = R_{(Info)} / (R_{(Info)} + R_{(NoInfo)})$. Values of $L_{(Info)} < 0.5$ or $L_{(Info)} > 0.5$ indicate a preference for *Info* or *NoInfo* respectively, as measured in forced trials, independently of the measure of preference based on choices in 2-option trials.

Ethical note

All experiments were carried out in compliance with the UK Animal (Scientific Procedures) Act (1986) and its associated guidelines.

Results

Training

Cue discrimination

A condition for the interpretation of preferences is that subjects were able to discriminate the contingencies of each cue; we examined this using cumulative head-poking time during the 10s interval preceding reward outcomes, when the cues were present, pooling data from the last three training sessions across the groups, which up to that point had no differential experience (Fig. 2). The cue associated with 100% reward probability (S+) had the longest cumulative head poking duration (2.98s ± 0.14; mean± s.e.m.), followed by the mean of both cues associated with 50% probability (N1 & N2: 2.40s ± 0.13), while the cue associated with no reward (S-) elicited the shortest average head poking duration (1.19s ± 0.09). A one-way repeated-measures ANOVA revealed a significant effect of cue ($F_{2,46} = 44.38$, $P < 0.0001$). Post-hoc pair-wise comparisons showed a significant difference between all pairs (100% vs 50%, $P < 0.05$; 100% vs 0%, $P < 0.001$ and 50% vs 0%, $P < 0.001$). This confirms that subjects discriminated the contingencies programmed for each cue.

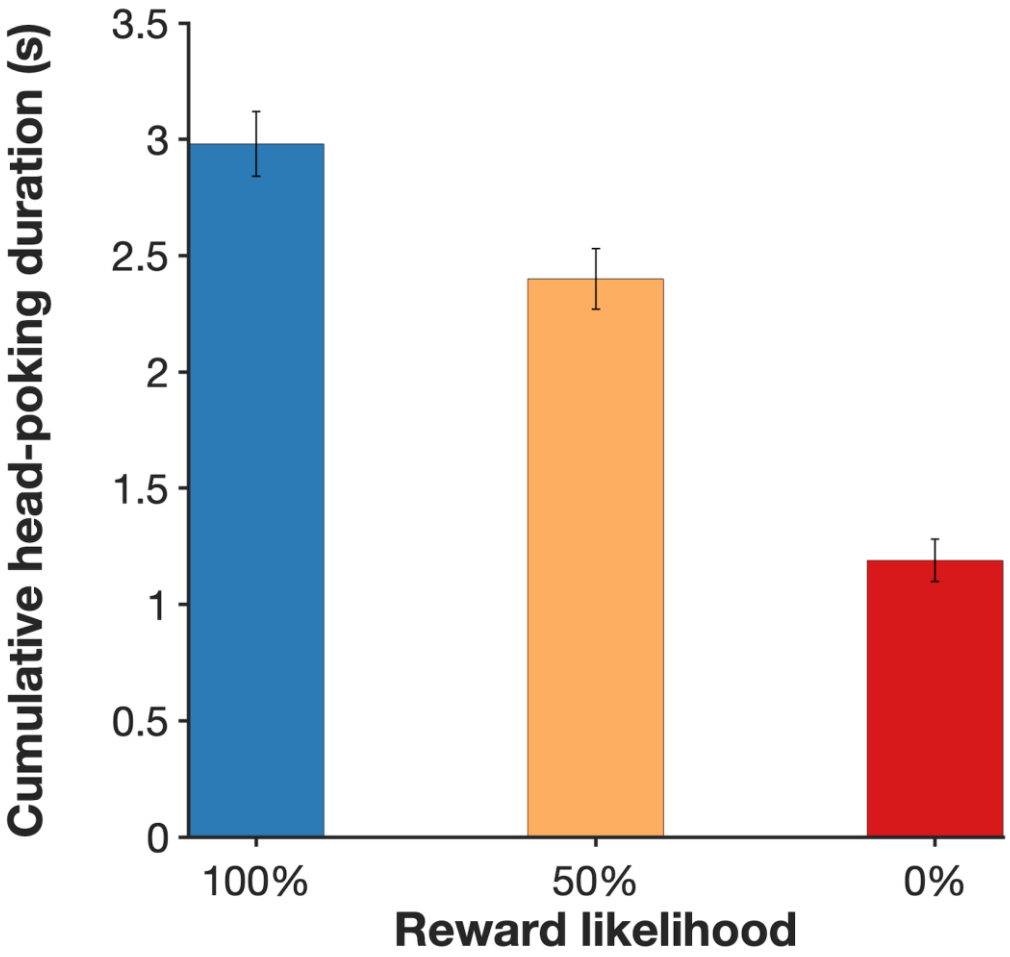


Fig. 2 Time spent head-poking into the food magazine during cue presentation at the end of the training phase. The data shows the mean cumulative time (mean ± s.e.m.) subjects spent with their head in the food magazine over the 10s intervals preceding reward outcomes, pooled from the last three sessions of training. During this time reward-predictive signals indicating a 100% (blue), 50% (yellow) or 0% (red) chance of reward were presented (corresponding to S^+, N1 or N2, and S^-, respectively). n = 24.

Experiment

Preference 1: Choice in 2-option trials

In choice trials a strong preference for *Info* developed in all three treatments, with acquisition occurring more slowly in the *Only_S*$^-$ treatment (Fig. 3). A two-way repeated-measures ANOVA on data across all sessions with treatment as a between-subject factor, session as a within-subject factor, and (transformed) proportion of choices for *Info* as the response variable, revealed significant effects of treatment ($F_{2,21} = 4.00$, $P < 0.05$) and session ($F_{13,276} = 14.4$, $P < 0.0001$), and a significant interaction ($F_{26,273} = 1.56$, $P < 0.05$), reflecting the slower acquisition in the *Only_S*$^-$ treatment. Given the significant interaction, and the plot in Fig. 3, it is obvious that the main effects are caused by rate of acquisition and not by asymptotic levels.

To establish whether the acquired preferences for *Info* were stable rather than transient we analysed preferences at the end of the experiment by pooling data over the last three sessions. In all three treatments the animals showed a strong preference for the informative option for as long as they were tested: 99.8% ± 0.002 (mean± s.e.m.) in the *S*$^+$*_S*$^-$ treatment*;* 90% ± 0.030 in *Only_S*$^-$*,* and 98.8% ± 0.005 in *Only_ S*$^+$. These values are all significantly greater than 50% ($t_7 = 47.5$, $P < 0.0001$; $t_7 = 5.86$, $P < 0.001$; and $t_7 = 22.8$, $P < 0.0001$, respectively).

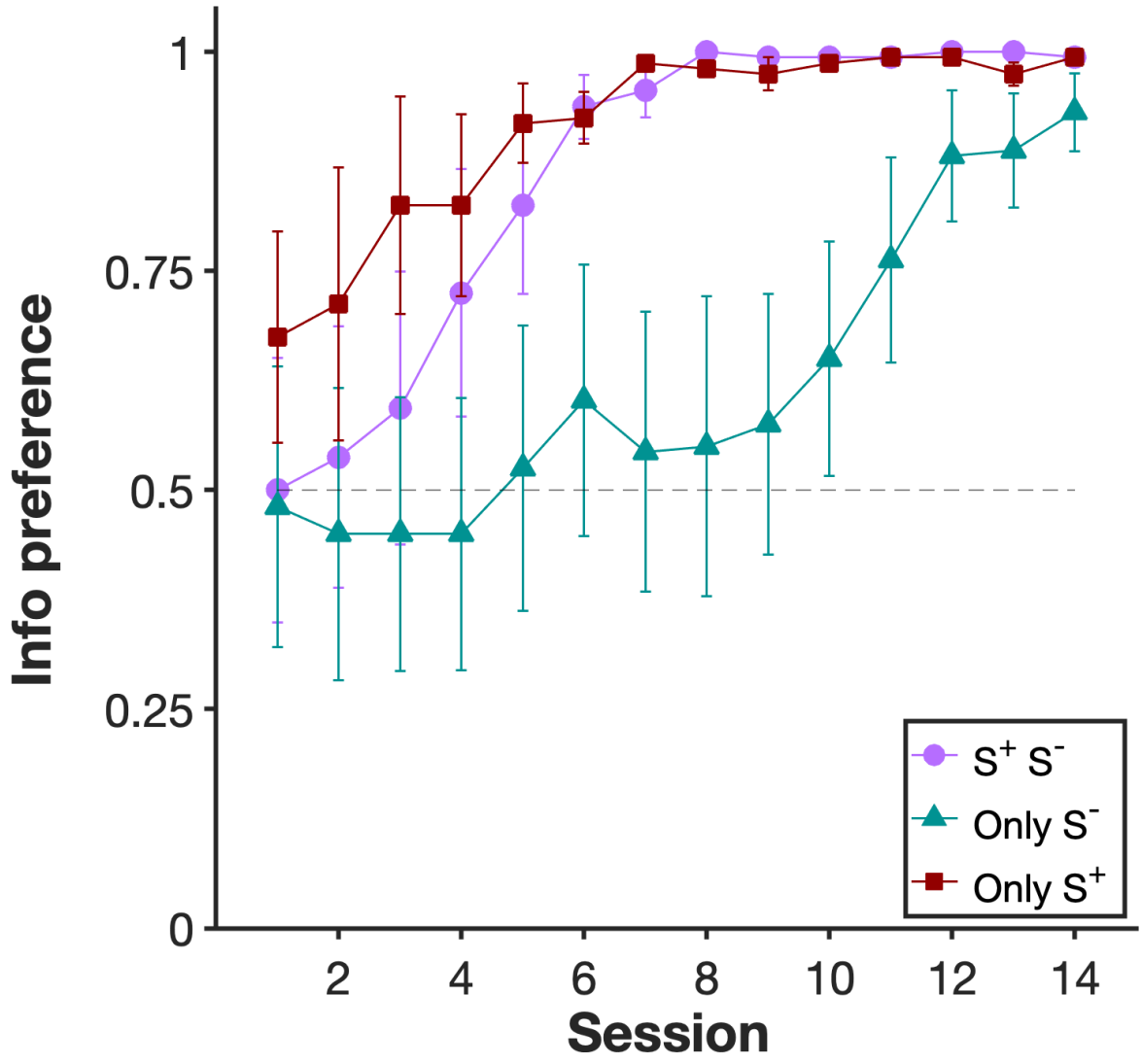


Fig. 3 Preference for the *Info* option in choice (2-option) trials. Proportion of observed choices for the $S^+_S^-$ (n = 8), $Only_S^-$ (n = 8) and $Only_S^+$ (n = 8) groups are shown (means ± s.e.m.).

Preference 2: Latency in 1-option trials

In the previous section we measure preference using proportion of choices in trials when both alternatives were present, here we use latency to respond in

single-option forced trials. This is the time between a subject initiating a trial by pressing the back lever and pressing the lever for the single option that subsequently becomes available. Latencies, or reaction times in 1-option trials have proven to be a robust predictor of choice in 2-option trials, and are very informative with respect to the psychological mechanism of choice (see, for instance, Monteiro et al., 2020). Since in each session each individual completed 20 *Info* and 20 *NoInfo* forced trials, we used the median latency shown by each individual for each alternative for analysis. Fig. 4 a) shows that latencies in single-option trials mirrored the rats' choice proportions in choice trials: in all treatments, latencies were shorter in *Info* than *NoInfo* in the final sessions of the experiment. The absolute value of latencies is shown in Fig. 4 b) and reveals that while latency to respond to *Info* was fairly constant across treatments, latency towards *NoInfo* varied: it was very long in the $S^+_S^-$ treatment, intermediate in *Only_S*$^-$ and shortest in *Only_S*$^+$. This is interesting because *NoInfo* was identically programmed across treatments; we return to this point in the discussion.

To quantify the acquisition of preference using forced trials data we ran a two-way repeated-measures ANOVA on data across all sessions, with treatment as a between-subject factor, session as a within-subject factor and latency-based preference index $L_{(Info)}$ as the dependent variable (see Methods). This revealed a

significant effect of treatment ($F_{2,21} = 9.72$, $P < 0.01$), session ($F_{13,273} = 17.54$, $P < 0.0001$), and a significant interaction ($F_{26,273} = 2.97$, $P < 0.0001$).

Post-hoc pair-wise comparisons on data pooled from the last 3 sessions showed that while $L_{(Info)}$ in *Only_S*$^-$ (0.43± 0.01 mean± s.e.m.) and *Only_S*$^+$ (0.43± 0.01) were not significantly different from each other ($P = 1$), $L_{(Info)}$ in both of these groups was significantly higher compared to the S^+_S^- group (0.33± 0.01; $P <$ 0.0001 in both cases). Further, consistently with a preference for *Info*, over the last 3 sessions $L_{(Info)}$ was significantly lower than 50% in all treatments (S^+_S^-: $t_7 = -7.02$, $P < 0.001$; *Only_S*$^-$: $t_7 = -3.74$ $P < 0.01$; *Only_S*$^+$: $t_7 = -4.16$, $P < 0.01$), indicating that on this metric the subjects also preferred *Info*. In other words, according to this index, preference for *Info* persisted at the end of training, and was strongest when both outcomes were explicitly signalled, and equally strong when either of the outcome signals was omitted.

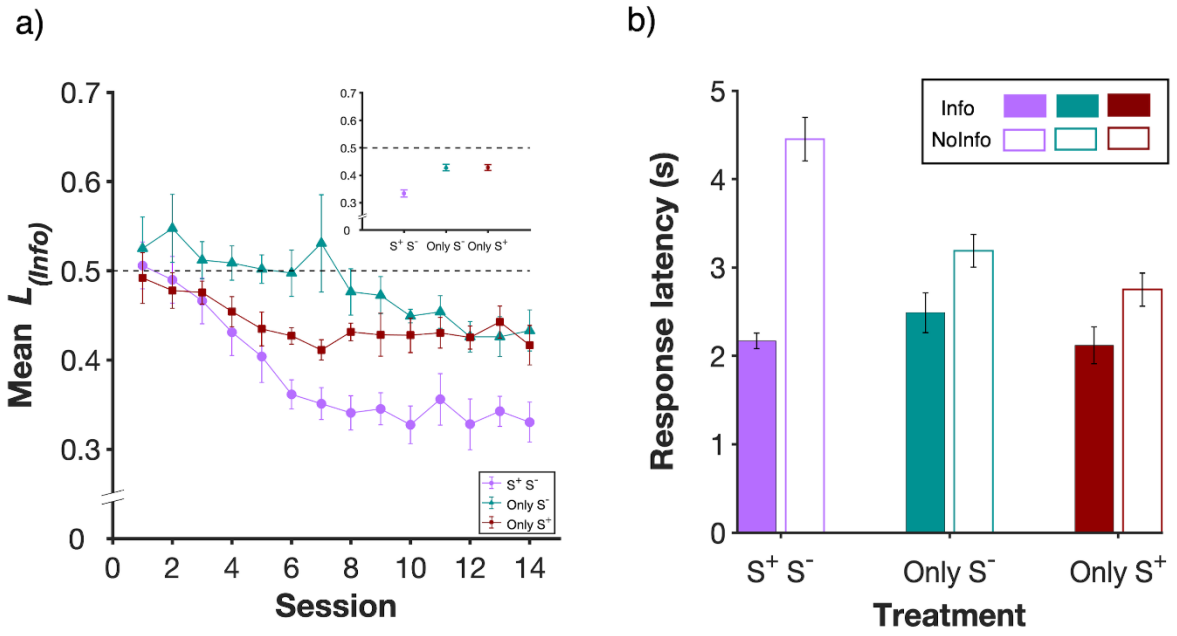


Fig. 4 Latency to respond to *Info* vs *NoInfo* in forced (1-option) trials. a) Latency-based preference index for all three treatments where $L_{(info)} = R_{(Info)} / (R_{(Info)} + R_{(NoInfo)})$, and $R_{(Info)}$ and $R_{(NoInfo)}$ are the median latencies to respond in *Info* and *NoInfo* respectively. $L_{(Info)}$ values below 0.5 indicate preference for *Info* while values of $L_{(Info)}$ above 0.5 indicate preference for *NoInfo*. The inset shows data pooled from the over the last 3 sessions. n = 8 in each group. **b)** Filled bars show latency to respond to *Info* ($R_{(Info)}$) and unfilled bars show latency to respond to *NoInfo* ($R_{(NoInfo)}$) across the three treatments (means ± s.e.m.), with data pooled from the last three sessions. n = 8 in each group.

Head-poking during cue presentation

Although during the main experiment behaviour post-choice did not influence outcomes, rats anticipated food by head-poking into the food magazine (possibly a Pavlovian response). Data from choice and forced trials show that in the *Info* option, subjects head-poked more in trials when food delivery was due than when it was not and showed an intermediate level of head-poking in *NoInfo*, when there was a 50% chance of food delivery **(**Fig. 5**)**.

Pooled over the last three sessions, time spent head-poking into the food magazine ranked as reward probability (100% > 50% > 0%). This was the case in all treatments: the $S^+_S^-$ group (100%: 0.83s ± 0.13; mean ± s.e.m., 50%: 0.48s ± 0.1, 0%: 0.19s ± 0.04), *Only_S*$^-$ (100%: 1.76s ± 0.21, 50%: 0.87s ± 0.16, 0%: 0.70s ± 0.12), and *Only_S*$^+$ (100%: 1.32s ± 0.24, 50%: 0.75s ± 0.12, 0%: 0.34s ± 0.06).

A two-way ANOVA on these data with reward probability as a within-subject factor, treatment as a between-subject factor and cumulative head-poking as the response variable revealed a significant effect of reward probability (100%, 50% or 0% reward; $F_{2,42} = 27.80$, $P < 0.0001$) but not treatment ($F_{2,21} = 2.53$, $P = 0.104$) and no significant interaction ($F_{4,42} = 0.51$, $P = 0.732$). Post-hoc pair-wise comparisons showed that head-poking was significantly higher when reward was

due (100%) than when it was not (0%) in all treatments (highest $P < 0.001$). The fact that head-poking reflected forthcoming reward outcomes differentially in S^+ and S^- trials regardless of treatment shows that rats recognized the current contingency regardless of whether an explicit cue was present. The absolute level of head-poking seemed to be inversely related to how much food signalling was available (*Only_S*$^-$ > *Only_S*$^+$ > *S*$^+$*_S*$^-$), as if attention to explicit signalling competed with exploratory investigation of the food magazine, though this is a non-significant trend.

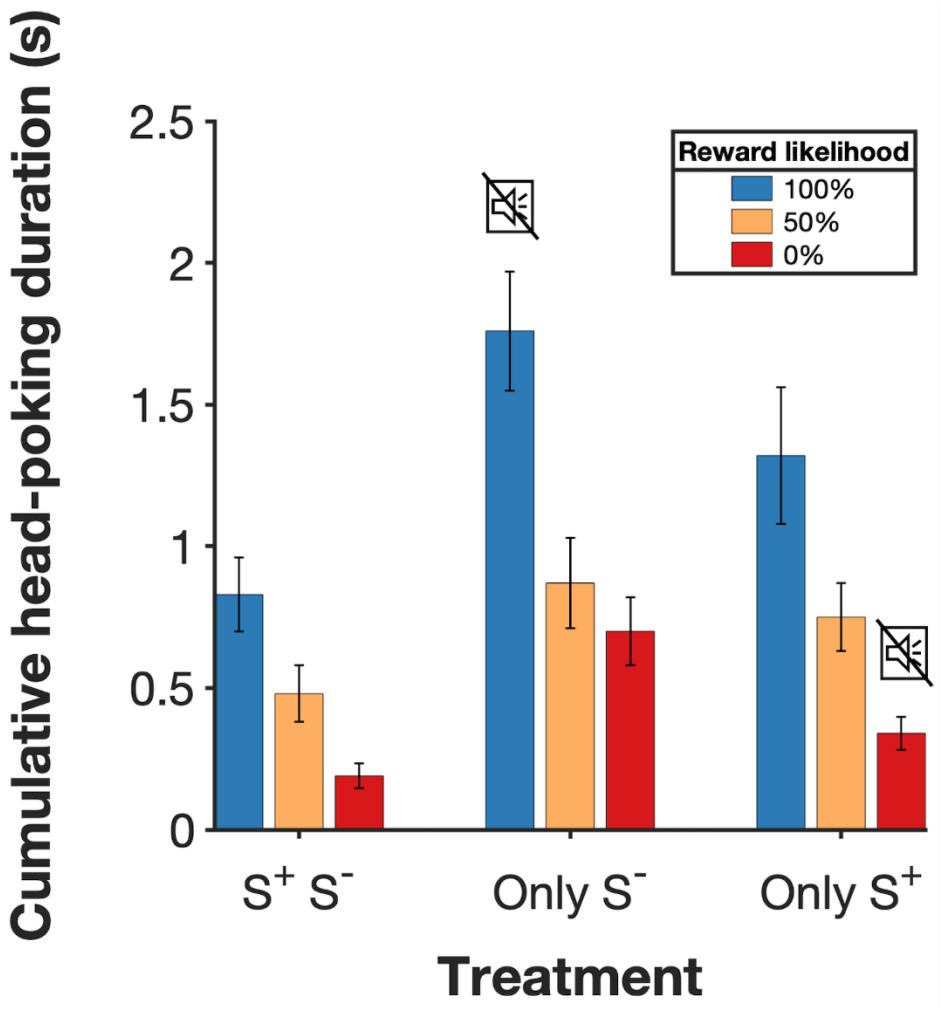


Fig. 5 Time spent head-poking into the food magazine between choice and outcome in the main experiment. The bars show the average cumulative time subjects spent with their head in the food magazine in the 10s preceding reward outcomes (± s.e.m.), pooled over the last 3 sessions. During this time reward-predictive signals indicating a 100%, 50% or 0% chance of reward could be presented. Data for the $S^{+}_S^{-}$ (n=8), *Only_S*$^{-}$ (n=8) and *Only_S*$^{+}$ (n= 8) groups are shown. The muted speaker symbol indicates that an explicit cue was not used to signal a particular outcome.

Discussion

We explored the role that two putative psychological mechanisms – uncertainty reduction and conditioned reinforcement - have in determining preference for an informative option in which delayed outcomes are signalled by predictive cues, over an equally profitable non-informative option, in which outcomes remain uncertain until they are realised. A pre-existing observation that is considered to be functionally paradoxical and mechanistically unclear, is that in such protocols animals show a strong bias for the informative option, even though the information they gain is non-instrumental, i.e., cannot be used to modify outcomes and increase rewards. As an aside, note that we label this preference as being paradoxical (which relates to the observer's expectations) and not, as it is frequently done, 'suboptimal' (see Zentall, 2016) which could be interpreted to suggest that that the mechanism generating the behaviour is maladaptive, a possibility which seems unlikely given the prevalence of preference for advanced information in a range of distantly related species (Ojeda et al., 2018).

We relied on two independent metrics of preference: proportion of choices in 2-option trials, and response latency in 1-option trials. As we show below, this helps to judge the robustness of preferences and to unravel behavioural mechanisms. In the $S^{+}_S^{-}$ treatment, where we reproduced the classic 'paradoxical choice'

protocol, our results are consistent with previous studies in rats: when presented with two options that differ only in the post-choice predictability of delayed outcomes, rats (as birds and primates) strongly prefer the more informative alternative (Chow et al., 2017; Cunningham and Shahan, 2019; Ojeda et al., 2018). This was observed both in proportion of choices between the alternatives and in differential response latencies when only one of them was present. Some previous studies on rats failed to find preference for the informative option (Alba et al., 2018; López et al., 2018; Martínez et al., 2017; Trujano et al., 2016; Trujano and Orduña, 2015), though this is likely because in these studies *Info* had a lower probability of reward than *NoInfo*, whereas in our study both options were equally profitable, and unlike pigeons and starlings, rats are very sensitive to reward losses incurred by selecting *Info* (Fortes et al., 2016; Ojeda et al., 2018; Smith et al., 2016; Vasconcelos et al., 2015).

Our results show that asymptotic *Info* preference is robust to the absence of an explicit, salient good news (*Only_S*$^{-}$ treatment), or bad news (*Only_S*$^{+}$ treatment) stimulus (Fig. 3). When the period preceding reward or no reward respectively was filled with silence rather than an auditory cue, subjects still developed a strong preference for *Info* in both cases. Our finding that a salient S^{+} is not necessary for *Info* preference is consistent with similar observations in starlings (Vasconcelos et al., 2015), monkeys (Lieberman, 1972) and humans (Fantino and

Silberberg, 2010; Lieberman et al., 1997). Fig. 3 does show however that the absence of a salient S^+ in *Only_S*$^-$ slows the acquisition of *Info* preference relative to the standard *S*$^+$*_S*$^-$ treatment, while the absence of a salient S^- in *Only_S*$^+$ has a much weaker, although positive, effect on the speed of acquisition. This result is congruous with several studies indicating that S^+ has a more significant impact on preference in paradoxical choice than S^- (Fortes et al., 2017; Laude et al., 2014; McDevitt et al., 1997; Pisklak et al., 2015; Spetch et al., 1994) and also those showing that in rats S^- acquires inhibitory properties (Alba et al., 2018; Martínez et al., 2017; Trujano et al., 2016; see González and Blaisdell, 2021 for evidence of this in pigeons).

At the beginning of the experiment, we found a number of subjects across all treatments showed extreme choice proportion values, even though they hadn't been exposed to option contingencies for long. This extreme inter-individual variation likely reflects side-biases that individual subjects have when initially presented simultaneously with two alternatives. Such biases are a well-documented phenomenon in 2-alternative-forced-choice tasks such as paradoxical choice, reflecting that subjects can express stereotyped and repetitive behaviour toward experimental manipulanda prior to much experience with the option contingencies. In the design of this experiment the side of both available alternatives was fixed for each subject but counterbalanced across subjects. This

means that these side-biases will be reflected in the choice proportions of individuals. As some subjects develop biases towards the right choice lever and others to the left this will be recorded as a very high or very low proportion of choices for Info depending on the side Info is assigned to for each subject. Counterbalancing the side of the informative option across subjects helps to ensure that such biases do no systematically bias average choice proportion data. Forced trials also help to mitigate the effect of side biases by ensuring subjects with experience the contingencies of both alternatives despite such biases.

Our analysis of preference on the basis of latency in single option trials is inspired by the Sequential Choice Model (SCM; Kacelnik et al., 2011; Monteiro et al., 2020; Shapiro et al., 2008). The SCM postulates that choice can be modelled as a horserace between the latency distributions of available alternatives because the alternatives are psychologically processed in parallel, without an active process of choice. Measuring behaviour by more than one procedure is in itself important, because if the phenomenon being measured is meaningful, it should show procedural invariance, a property often claimed to be violated by studies of human preferences using choice Vs. willingness to pay (Slovic, 1995). We did find general, qualitative consistency between our measures of preference, but also found that using response latency as an additional metric informed about important aspects of potential underlying mechanisms. As Fig. 4b shows, response latencies in forced trials for *Info* were consistently shorter than in trials

for *NoInfo*, across treatments. Variations between treatments were mediated only by latency differences in *NoInfo*, which was identically programmed in all three treatments. In other words, treatment effects were mediated by modifications of latency to respond to the least preferred alternative. This result is striking, could not have been anticipated by the choice results, and is consistent with what was reported by Smith et al. (2018) in a midsession reversal protocol with pigeons, a very different experiment, and species. They too, found that changes in choice proportions were explained by variations in latency towards the least preferred alternative in single option trials, when that option did not itself change in its properties. It seems appropriate to infer that parallel processing of alternatives, and mediation through latency variation in less preferred alternatives can be widespread properties of choice behaviour, something that the analysis of choice, which is prevailing in studies of preference, could not have revealed.

We also examined differences in response latency between Info and NoInfo across treatments over the course of the main experiment using a latency index, where lower values indicate faster responses towards Info relative to NoInfo. While subjects did respond more quickly towards Info across treatments, the dynamics of relative differences in response times did not reflect the dynamics of preference acquisition in choice across the three groups. Specifically, quicker responding towards Info developed rapidly in $S^+_S^-$ while in *Only_*S^- and *Only_*S^+ faster responses towards Info developed more slowly, and in the terminal phase

of the experiment the latency index was lowest in $S^+_S^-$ while similar across *Only_S*$^-$ and *Only_S*$^+$ (Figure 4a). This pattern is inconsistent with preference acquisition in choice trials where acquisition is most rapid in *Only_S*$^+$ and $S^+_S^-$ and significantly slower in *Only_S*$^-$. It is not entirely clear why this discrepancy between both metrics of preference exists, but it is perhaps notable that the pattern of results found in the latency index where *Only_S*$^-$ and *Only_S*$^+$ share more similarity than $S^+_S^-$, can also be found (though without statistical significance) in head-poking data from the end of the experiment: head-poking during queue presentation is generally elevated in the treatments where explicit reward signalling is lower frequency compared to $S^+_S^-$. It is therefore possible that as with head-poking, but not preference in choice, response latencies reflect greater levels of perceived ambiguity in outcomes owing to omissions of explicit reward signals, and that this may account for quantitative discrepancies between both metrics of preference.

We focused on two potential psychological hypotheses about the mechanisms supporting the observed bias for the informative option. The information hypothesis contends that individuals treat uncertainty as aversive, so that informative signals, regardless of whether they bring good or bad news, drive preference acquisition. In contrast, the conditioned reinforcement hypothesis argues that preference for the *Info* option increases due to signals for food (S^+, 'good news') and decreases due to signals for food's absence (S^-, 'bad news'),

because S^+ acquires secondary excitatory properties, and S^- inhibitory properties, with the excitatory influence of S^+ deemed to be greater than the inhibitory effect of S^-.

Though both mechanisms are plausible, both run into functional difficulties. For the information hypothesis this is that in the experimental situation, acquiring information does not give subjects the ability to increase reward outcomes (but see Dinsmoor, 1983 for the suggestion that it may improve the usefulness of outcomes). This difficulty, like other experimental observations of so-called suboptimal or irrational behaviour, can be addressed post-hoc by arguing that in nature, information about relevant commodities is very often likely to be usable, either immediately or further in the future, so that evolution may design utility functions that are somehow tricked by the experimental protocols. For example, foraging-inspired theoretical models (Freidin and Kacelnik, 2011; Vasconcelos et al., 2015) have argued that in nature, information, even if it announces unfavourable events, is likely to be immediately useful: an animal that knows for sure that the prey being presently pursued will not be captured, would abort the chase, and thus would not pay the opportunity cost of waiting for a null outcome. Thus, Vasconcelos et al. (2015) have argued that in the lab animals do not include the delays associated with S^- in their profitability computations, because they would normally use such time beneficially. Even if the information cannot be

used straight away, acquiring it may help to solve novel problems in the future (Gottlieb and Oudeyer, 2018). In other words, it is the artificiality of being unable to use information in the experimental protocol that generates the paradox, which can be reconciled by considering the ecological context in which the mechanism of behaviour evolved (Vasconcelos et al., 2018). Similarly, with the conditioned reinforcement hypothesis, while there are no *a priori* reasons why the excitatory effect of S^+ should be greater than the inhibitory effect of S^- it is likely that in nature cues indicating the presence of relevant commodities are more prevalent or reliable than those indicating their absence, therefore the power of excitatory and inhibitory conditioned stimuli to modify behaviour need not be symmetric (e.g., Rescorla & Wagner, 1972). Placing the phenomenon in its natural ecological circumstances may be key to promote its understanding.

Our results do not lend unequivocal, exclusive support to either the conditioned reinforcement hypothesis or the information hypothesis in their original formulations. However, an extended conditioned reinforcement account (which we detail below), or both mechanisms acting simultaneously, may explain our results. We discuss these putative explanations below, focussing first on the information hypothesis.

The information hypothesis makes two predictions that distinguish it from the conditioned reinforcement account. The first is that S^- on its own should reinforce

Info preference. Our main result is consistent with this prediction: in the *Only_S*$^{-}$ treatment group where a salient S^{+} was absent, but S^{-} present, rats also acquired a strong preference for *Info,* which could be interpreted to show that S^{-} reinforces *Info* responses, rather than just inhibiting them. Head-poking data from the delay period between choice and reward outcomes, however, could be taken to suggest otherwise. Across all treatments we found that time spent head-poking into the food magazine ranked as reward probability (100% > 50% > 0%; Fig.5). In other words, when an explicit S^{-} cue was present, magazine head-poking was lower than when S^{+} or an uninformative cue were present, indicating that S^{-} may have inhibitory properties. Note though that the fact that S^{-} may inhibit post-choice head-poking does not necessarily preclude it from reinforcing *Info* choice (which occurs earlier in the trial) via uncertainty reduction.

A quantitative prediction of the information hypothesis is that S^{+} and S^{-} should reinforce choices for *Info* to the same extent. This is because both stimuli in the informative option completely resolve the pre-choice uncertainty, hence they convey the same amount of information. Thus, if uncertainty reduction were the only consideration, S^{+} and S^{-} would be equally reinforcing. Our finding that the omission of an explicit S^{+} slows acquisition to a greater extent than the omission

of an explicit S^- is incongruent with this prediction. Taken together therefore, our results cannot be fully explained by the information hypothesis alone.

According to the conditioned reinforcement account, animals prefer the informative option because of the excess excitatory effect of good news. Explanations of precisely how S^+ can acquire value as a conditioned reinforcer in paradoxical choice have been developed by several different authors and include: the Contrast Hypothesis (Case and Zentall, 2018; Gipson et al., 2009; Zentall, 2013; see also González et al., 2020 for a hypothesis that considers contrast but not conditioned reinforcement *per se*), the Stimulus Value Hypothesis (Smith et al., 2016; Zentall et al., 2015; Smith and Zentall, 2016), the Signals for Good News (SiGN) Hypothesis (Dunn and Spetch, 1990; McDevitt et al., 2016), the Temporal Information Model (Cunningham and Shahan, 2018 though note that their model also considers how primary reinforcement affects choice), and the Selective Engagement Hypothesis (Beierholm and Dayan, 2010; Dinsmoor, 1983).

We do not have the scope or the data to examine and differentiate these hypotheses in detail, but they all share the assumptions that (1) S^+ alone is responsible for acquisition and maintenance of *Info* preference, and (2) the excitatory effect of S^+ is greater than the inhibitory effect of S^- (with some

claiming that S^- has no effect at all). The conditioned reinforcement account on its own therefore predicts that preference for *Info* can only develop when there is a perceivable S^+ cue in *Info* capable of acquiring reinforcing properties.

The fact that subjects in the *Only_S*$^-$ treatment, where there was no explicit auditory S^+ signal, developed preference for *Info*, *prima facie* contradicts the conditioned reinforcement account, as the result suggests that S^+ is not necessary for the development of *Info* preference. However, an elaborated conditioned reinforcement account could explain our results, and we discuss this below.

It could be argued that the manipulation we performed was not sufficient to eliminate the putative positive conditioned reinforcement afforded by the informative option. After all, head poking data showed that during the post-choice delay rats could anticipate whether food was imminent, even for outcomes not signalled by a salient cue (Fig. 5). A conditioning explanation for this could be that subjects treat the compound of their action (lever pressing) plus the immediate absence of a salient cue as a predictive event or conditioned stimulus (CS) in itself, and the delayed outcome (food or no food) as the unconditioned stimulus (US). Thus, they could learn the pairing *[Press Info+Silence]* ➔ *food* in the *Only_S*$^-$ treatment, while those in *Only_S*$^+$ could learn *[Press Info+Silence]* ➔ *no food*. A Pavlovian version of the same idea is that the CS

compound does not comprise just the rat's action, but also the lever retraction that follows from it. It is therefore possible that in the *Only_S*$^{-}$ group, *Info* lever pressing/retraction followed by the *absence* of an auditory cue is a compound stimulus used by rats to anticipate reward, in other words, it is a virtual S^{+}. Under this rationale, conditioned reinforcement can be present in *Info* even with no salient perceptual cue precedes rewards, and thus could account for the development of preference for *Info* in our experiment.

Paired with another well-established psychological phenomenon, the feature-positive effect (Crowell and Bernhardt, 1979; Newman et al., 1980; Pace et al., 1980; Sainsbury, 1971), this elaboration of the conditioned reinforcement hypothesis could provide a full explanation of both preference and speed of acquisition across our treatments. The feature-positive effect refers to the observation that in discrimination learning involving the presence or absence of a feature, subjects find it easier to associate the presence of a feature with a positive outcome (S^{+}) than they do the absence of the same feature. This could explain why subjects developed *Info* preference faster in *Only_S*$^{+}$ compared with *Only_S*$^{-}$. In *Only_S*$^{-}$ food rewards in *Info* are preceded by the lack of an auditory cue; a feature negative stimulus, while in *Only_S*$^{+}$ food is preceded by an explicit, feature-positive, auditory cue. Therefore, the feature negative silence in *Only_S*$^{-}$ would have taken longer to acquire secondary reinforcing properties than the

feature positive reward cue in *Only_S*$^{+}$, which may explain the difference in acquisition between the treatments. This could be because one of the components of the potential compound cue for reward in *Only_S*$^{-}$ (*[Press Info+Silence])* was silence (or background noise), which itself is associated with the absence of reward during the ITI. Therefore, we cannot rule out conditioned reinforcement as the primary underlying mechanism driving *Info* preference in our experiment.

Finally, our results are also consistent with the possibility that both uncertainty reduction and conditioned reinforcement act simultaneously to generate preference in paradoxical choice (see Daddaoua et al., 2016 for a similar argument in monkeys). An asymptotic preference for *Info* that is robust to the absence of an explicit S^{+} cue is consistent with the information hypothesis prediction that a perceivable S^{-} alone is sufficient to generate *Info* preference via uncertainty reduction. Additionally, the faster acquisition in *Only_S*$^{+}$ where there is an explicit S^{+} compared to *Only_S*$^{-}$ where there is not, supports the conditioned reinforcement hypothesis' assumption that a perceivable S^{+} reinforces *Info* choices. Taken together intrinsic information value and conditioned reinforcement can therefore provide a reasonable account of our results and those of other studies. This mechanism would capture as significant two functionally relevant commodities - both the *amount* of information and its *content* (i.e., good

news or bad news) – as factors that shape the acquisition of preferences in the paradoxical choice protocol.

In summary, we found that rats show a robust preference for advanced non-instrumental information and that this preference is more strongly influenced by good news than bad news. Counterintuitively, treatment effects were mediated by differences in the latency to respond to the least preferred alternative, which was identical across all treatments. Our results show that while uncertainty reduction is unlikely to solely account for preferences for advanced information, the balance of evidence indicates that it may play some role alongside conditioned reinforcement.

Supplementary materials

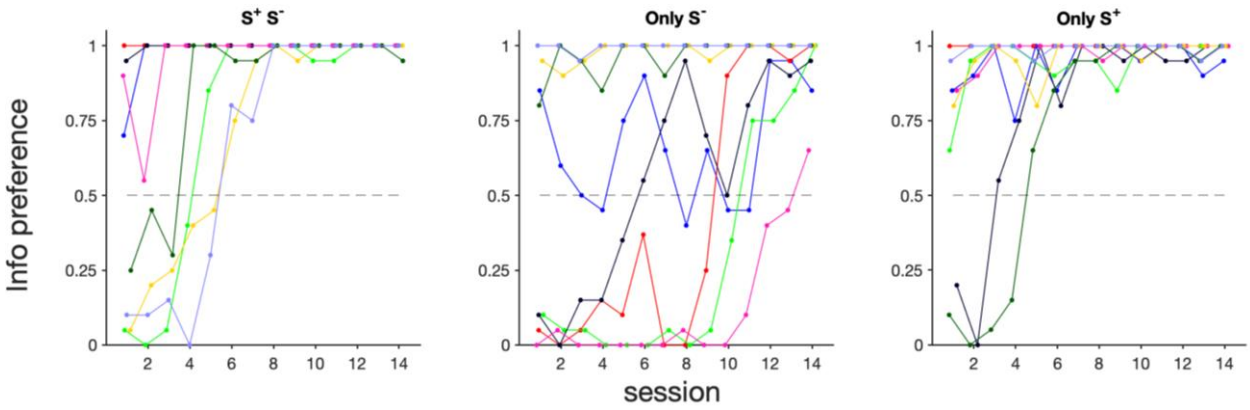


Figure S1. Individual preferences for the *Info* option. Proportion of observed preferences in the $S^+_S^-$ (n = 8), *Only*_S^- (n = 8) and *Only*_S^+ (n = 8) groups are shown. Each colour represents a different individual for each group.

Author List: Victor Ajuwon, Andres Ojeda, Robin A. Murphy, Tiago Monteiro and Alex Kacelnik

Author contributions: VA, AK, AO and RAM conceptualised and designed the experiment; VA and AO collected the data; VA, AO, TM, and AK analysed the data; VA and AK wrote the first draft of the paper; all authors edited and reviewed the manuscript.

Chapter 2. Non-instrumental information promotes risk-proneness for amounts of food in rats

Abstract

Real-life decision-makers can respond to stochasticity through preferences based on the experienced variance of alternatives (risky choice), or by seeking opportunities to gain information. The interaction between risky choice and information seeking is not well understood, and while research suggests that animals are mainly risk averse for amounts of reward, it is also known that they would afford lower gain rates just to reduce uncertainty, even when this does not improve gains in the long term. We investigate this interaction with rats as experimental subjects. Across three treatment groups, we examined rats' preferences between alternatives of equal profitability (expected food / expected time) that differed only in a) the availability of advanced information about stochastic reward amounts; *Risky(Info) vs Risky(NoInfo*) b) risk in amount of reward; *Risky(NoInfo) vs Fixed,* and c) both the availability of advanced information and risk; *Risky(Info) vs Fixed*. Consistently with previous work on 'paradoxical choice', in a) rats preferred advanced information about the magnitude of forthcoming rewards even though the information could not be used to improve returns. Consistently with research on risky choice, in b) subjects showed risk aversion for amounts of reward. However, when, in c), the risky alternative also offered advanced (i.e., post-choice) information about forthcoming outcomes, subjects were risk prone, preferring the option yielding

variable outcomes. These results show that non-instrumental information can modulate preferences for risk. The difference in risk-related preferences cannot be readily accounted either by Risk Sensitivity Theory, which predicts that shifts in risk proneness should track the energetic state of decision-makers, or by Scalar Utility Theory, that uses psychophysics to predict risk aversion for positive values such as food amounts. Our results highlight how critical it is to combine functional and mechanistic modelling and experimentation to unravel properties of decision making, including risk sensitivity.

Introduction

Evolutionarily-inspired models of animal behaviour posit that because natural selection has acted on behavioural mechanisms, organisms appear designed to maximise their reproductive success, or inclusive fitness (Grafen, 2009; Parker and Smith, 1990). In order to achieve this when foraging for food, a key assumption has been that, all else being equal, individuals should prefer options that maximise their net long-term rate of energy intake (MacArthur and Pianka, 1966), and be indifferent between alternatives with the same average pay-off. This is a sensible assumption given that this strategy would minimise opportunity costs in the form of time, and maximise energy intake - both outcomes that should result in greater effort being directed to reproduction. This assumption therefore underlies a range of influential models in the foraging literature (Stephens and Krebs, 1986) e.g., Charnov's (1976) Marginal Value Theorem, and remains a prominent component of models of value-based decision-making.

However, in natural settings, foraging alternatives are highly likely to differ in reward properties other than average reward pay-off that nonetheless have significant impacts on survival and fitness. Therefore, to fully understand how agents optimise their choices to maximise fitness during foraging, we must consider how a range of other reward-related factors influence behaviour. In this

paper we investigate how two such distinct, but related properties of food sources interact to generate preferences. The first is risk, which can be defined as quantifiable stochasticity or unpredictability in reward outcomes pre-choice. As an illustration, an alternative that unpredictably alternates between a high and low reward amount is defined as risky while one that delivers a fixed pay-off is not (assuming the delays between rewards are fixed in both cases). The second property is informativeness, by which we mean how the degree of unpredictability or uncertainty changes in the course of time between choice and outcome (note that we use uncertainty and unpredictability interchangeably, rather than referring to Knightian uncertainty (Knight, 1921) which describes situations where probabilities and outcomes are entirely unknown). Both of these ideas are linked in the sense that only risky alternatives can also be informative, as information relates to reductions in unpredictability, and fixed alternatives are by definition already predictable. However, both riskiness and informativeness are distinct, as we illustrate in the sections below.

Though previous experiments have shown that both of these factors can influence foraging decisions (for reviews see Dunn et al., 2023; Kacelnik and El Mouden, 2013; McNamara and Houston, 1992; Zentall, 2016), they have largely been dealt with separately in different literatures, despite the fact that (1) agents will likely face trade-offs between avoiding risk and actively seeking information, and (2) advanced information about stochastic outcomes could affect the perceived and

experienced riskiness of a foraging option. Here, we address this by considering these factors jointly, presenting an experiment with rats that examines their interaction in choice. First, we introduce Risk Sensitivity Theory (RST), and then summarise several interpretations of the observation that animals appear to be intrinsically motivated to acquire advanced information about upcoming rewards, before considering the implications of the interaction between them.

Risk Sensitivity Theory

To illustrate the essence of RST, consider a forager that has access to two food sources repeatedly, before finally interrupting to roost overnight until the morning. Both alternatives offer the same average reward pay-off, but they differ in variance. The risky option unpredictably alternates between a high and a low pay-off, while the non-risky (fixed) option reliably delivers an intermediate amount of food. Now suppose that the forager has a minimum energy budget requirement: to avoid starvation overnight, it must have consumed some minimum amount of food by the end of the day. Which option should the forager prefer? The 'budget rule' (Stephens, 1981) provides an adaptive solution: if the intermediate pay-offs from the fixed option would take the forager above its minimum energy budget, then the forager should be risk averse and prefer the

fixed option. However, if the intermediate pay-offs from the fixed option are not sufficient, the forager should prefer the risky option, which in this case would at least offer the possibility of a string of high pay-off rewards that would keep it alive overnight. This basic idea was extended, among others by McNamara and Houston (1992), who incorporated a dynamic framing of the problem. This extreme scenario highlights that two alternatives that have the same average pay-off, but different variance, can have drastically different impacts on fitness.

More generally, the issue of preference between a variable and fixed option of equal average reward can be analysed by considering the function that relates the amount of food gained, x, to fitness value, $f(x)$, given that behaviour appears designed to maximise fitness. A mathematical result known as Jensen's inequality means that the shape of this function is of critical importance for determining preference between options that differ in variance. It states that if $f(x)$ is non-linear and x can take a range of values (as in the risky option), then the overall mean fitness gained - mean $f(x)$ - is not equal to f(mean x), hence the variable and fixed options will have different fitness payoffs. If the function is convex/accelerating then mean $f(x) > f$(mean x) and the variable option will result in higher fitness on average and be preferable. If the function is concave/decelerating, then the opposite is true. Following the logic of the budget rule, underbudget foragers should have convex fitness functions (be risk prone), while more satiated foragers should have concave ones (be risk averse).

Remarkably, Caraco et al. (1990, 1980) found evidence of this in yellow-eyed juncos. They inferred $f(x)$ across subjects at different levels of satiation by mapping the utility function subjects use to attribute value to reward gains, and observed the trends predicted by RST.

Over the years however, evidence for the budget rule has been sparse, with only a minority of studies on risk showing the expected shift to risk proneness when subjects are low on their energy budget (Kacelnik and El Mouden, 2013). A more striking trend is that independent of energy budget considerations, subjects tend to be risk averse for amounts of reward and strongly risk prone for delays to reward (e.g., Bateson and Kacelnik, 1995; Craft, 2016; for a review see Kacelnik and Bateson, 1996). From a functional perspective this is not entirely resolved, but McNamara (1996) has argued that in changing environments, risk averse behaviour for amounts of reward is likely to improve survival, particularly because learning about the properties of highly variable food sources is difficult, making them less attractive for foragers. From a mechanistic perspective, Scalar Utility Theory (SUT; Kacelnik and Brito e Abreu, 1998; Reboreda and Kacelnik, 1991) can account for risk aversion for amounts, by considering that amounts of rewards are perceived and remembered by animals with error. Weber's law (Foley and Bates, 2019) describes the phenomenon whereby the precision of perception decreases in proportion to stimulus value. When applied to rewards, SUT postulates that the error associated with an agent's estimate of amount

increases in proportion to magnitude such that in memory a given reward is represented as a bell-shaped probability-density function with mean and standard deviation proportional to the amount gained. When an alternative like the risky option yields two variable reward amounts, the two distributions are summed and the agent will remember that source as one continuous distribution. Because larger values are associated with greater variance, this summed distribution will be right skewed with a median lower that of the distribution of a fixed option delivering the same average amount of reward. Thus, the observation that agents are generally risk averse for amounts of reward, may arise from universal properties of perception, that in the context of risky foraging are beneficial in rapidly changing environments. The same rationale predicts risk proneness for variables whose value the agent prefers to be small, such as delay to food (Reboreda and Kacelnik, 1991).

Non-instrumental information-seeking

The consensus that animals are sensitive and even averse to the unpredictability associated with risky alternatives highlights that factors independent of reward rate can have significant impacts on value attribution and choice within foraging contexts. Recently, there has been a resurgence of interest in the effect that

informative reward predictive stimuli have on choice, even though such information cannot be used to improve gains in the lab.

In another experimental paradigm referred to as ‘paradoxical’ or ‘sub-optimal’ choice, subjects choose between two options that can each deliver a rich or poor outcome stochastically (e.g., small or large reward). In the informative option, a signal is immediately presented after the choice response, and this anticipates the trial outcome. In the uninformative option, there are no signals that allow subjects to predict the outcome ahead of time, so subjects are uncertain until the delay has lapsed. Critically, subjects cannot use the information gained in the informative option to increase returns, so the information is labelled ‘non-instrumental’. Intriguingly, when both alternatives are equally profitable i.e., have the same chance of giving a rich or poor outcome, rats (Ajuwon et al., 2022; Ojeda et al., 2018) and macaques (Bromberg-Martin and Hikosaka, 2009, 2011) show strong preference for the informative option. Unlike in risky foraging scenarios, this effect therefore cannot be attributed to a difference in variance between the options, though both are themselves risky at the time the animal makes its choice response. Remarkably, like humans performing similar tasks (Brydevall et al., 2018; Rodriguez Cabrero et al., 2019) a range of mammal and bird species (e.g., Blanchard et al., 2015; Fortes et al., 2016; Vasconcelos et al., 2015; Smith et al., 2016; White et al., 2019) show strong preferences for the informative option, even when doing so results in significant reward losses.

The paradoxical choice phenomenon is now well documented (for reviews see Cunningham and Shahan, 2018; Dunn et al., 2023; McDevitt et al., 2016; Zentall, 2016) and from a mechanistic perspective there are broadly two accounts. One interpretation is that these preferences reflect a primary reinforcing effect of information (the same argument has been made for similar protocols, see Dinsmoor 1983). Information's reinforcing properties may be derived from the fact that the resolution of uncertainty (Berlyne, 1960; Bromberg-Martin and Hikosaka, 2009; Hendry, 1969), or the closing of "information gaps" (Loewenstein, 1994) is itself appetitive, or because information enables agents to avoid temporally prolonged uncertainty, which is aversive (Bennett et al., 2016; Brydevall et al., 2018). The other account focuses more on the reinforcing properties of stimuli that announce positive outcomes, emphasising that stimuli for 'good news' are appetitive while stimuli for bad news are weakly inhibiting or neutral. Stimuli announcing positive outcomes may derive value through secondary conditioning by providing specific types of information about desirable outcomes (Cunningham and Shahan, 2018; Dunn et al., 2023; Smith et al., 2016), or by enabling subjects to savour positive outcomes while they wait (Ligaya et al., 2020, 2016). Furthermore, there is evidence that these two mechanistic accounts need not be mutually exclusive, but may guide behaviour by operating simultaneously (Ajuwon et al., 2022; Daddaoua et al., 2016).

From an adaptive perspective some authors have argued that the mechanisms driving these preferences are foraging-specific adaptations that are rate-maximising in natural scenarios when information about forthcoming food rewards is usable by foragers to modify behaviour appropriately (Fortes et al., 2017, 2016; Vasconcelos et al., 2018, 2015). Another interpretation is that preferences for an alternative providing advanced non-instrumental information in the lab are an analogue or homologue of human curiosity – the intrinsic desire to learn (Cervera et al., 2020; Gottlieb and Oudeyer, 2018). Therefore, the mechanisms driving them should function to enable agents to develop an accurate model of the world when operating in natural circumstances, ultimately helping them to solve a range of novel challenges that are not necessarily restricted to foraging.

Risk and Information

Though preferences pertaining to variance in stochastic sources (risk sensitivity) and information-seeking are both likely strategies that individuals deploy to adaptively deal with unpredictable variance in foraging contexts, little is known about the interaction between these modulators of preference and the underlying mechanisms. Furthermore, in specific versions of the paradoxical choice task where the informative option is associated with greater risk than the alternative

(e.g., Case and Zentall, 2018; Spetch et al., 1990; Zentall and Stagner, 2010), subjects have been shown to prefer the more informative but risky alternative, a finding that is incongruous with the consensus in the risk literature that subjects are generally averse to risk in reward amounts. This discrepancy warrants further investigation, highlighting that individuals may face trade-offs between uncertainty reduction and experienced stochasticity, an interaction that have received little theoretical and experimental investigation.

We address these issues empirically with rats (*Rattus norvegicus)* across three treatments. Choice alternatives differed either in informativeness *Risky(Info) vs Risky(NoInfo),* variance *Risky(NoInfo) vs Fixed* or in both informativeness and variance *Risky(Info) vs Fixed.* We designed the experiment such that the properties of each of the three possible pairs of options (an individual was only ever exposed to two) *Risky(Info), Risky(NoInfo)* and *Fixed*, were identically programmed in the treatments in which they were present, aiding intergroup comparisons of preference. In *Risky(Info) vs Risky(NoInfo)* we implemented a variation of the classic paradoxical choice protocol, giving subjects a choice between two options each with a 50% chance of delivering either a large (3 pellets) or small (1 pellet) reward after a 10s delay. Immediately after the choice response in the informative option signals anticipated which reward outcome was imminent, while subjects in the uninformative option were uncertain until after the delay. In *Risky(NoInfo) vs Fixed* we implemented a risk protocol with

Risky(NoInfo) unpredictably delivering 3 or 1 pellet on each trial (after a 10s delay) and the fixed option delivering a reliable 2 pellets on each trial (after a 10s delay). Finally, in *Risky(Info) vs Fixed* subjects could choose between a risky option whose outcomes were unpredictably variable (risky) pre-choice, but once selected revealed information pertaining to rewards, and the fixed option which was associated with no unpredictability, and therefore could not have informative signals. Some previous studies have examined the effect that reward predictive stimuli have on choice between a risky and fixed alternative (Case and Zentall, 2018; Zentall and Stagner, 2011), however in these cases baseline preferences for variability may have been masked by the fact that the risky alternative was also less profitable than the fixed one. Here, in each treatment group all alternatives are equally profitable, and in *Risky(Info) vs Risky(NoInfo),* both alternatives have equal outcome variance, allowing us to examine independently the effects of variance and advanced information, before determining how they interact on preference.

Methods

Subjects

24 male Lister Hooded rats (provider Envigo) < 12 weeks old at the start of the experiment served as subjects. They were housed in groups of four. Throughout the experiment they were food deprived to 85-90% of their free-feeding weight using growth curves from the provider. Water was provided ad-libitum in their home cages and they were maintained in a 12-hour dark/light cycle with lights on starting at 6 AM.

Apparatus

Testing was carried out in eight operant chambers (Med Associates, USA.) Each chamber contained three retractable levers: one in the back panel (centre) and two in the front panel, left and right of a central food magazine delivery tray equipped with an infrared beam and a sensor to record head entry and exit. Each reward delivery consisted of either three, two, or one 45mg sucrose pellet(s) (TestDiet, USA). A speaker was positioned above the food magazine in the front panel. Each chamber was also equipped with a house light (white) and a fan, which were switched on for the duration of the session. The chambers were controlled via

custom- written Med-State Notation programs running on MED-PC V (Med Associates, USA).

Cues

Four sounds, all with a duration of 10 seconds were used across the subjects as the cues: a low frequency pure tone (3 kHz, 78 dB), a high frequency pure tone (6 kHz, 78 dB), a buzzing sound (78 dB) and a clicking sound (74 dB). The combination contingencies associated with the four cues differed between the three treatments (see *Treatment Groups* and *Choice Alternative* sections below). Assignment of sounds to reward contingencies was (partially) counterbalanced across subjects within each group to avoid the possibility of option preferences being generated by any intrinsic aversive or attractive properties of the sounds.

Training

Magazine training

To train the rats to collect rewards in the apparatus, training began with a single session in which 5g of sucrose pellets were placed in the food magazine. Subjects were then placed in the operant boxes for an hour. In this session the house light and fan were not turned on. The second day of magazine training was a variable interval session where food was delivered on average once a minute (VI60 free food schedule) a total of 60 times. The variable interval was sampled from a truncated Poisson distribution with a mean of 60 s and range of 0-120 s.

Lever training

Over the next three sessions the rats were trained to press the two front levers. Either lever (left or right with equal probability) was available on each trial (60 trials per session). Once a lever extended into the chamber, a single press resulted in its retraction and immediate reward delivery (Fixed Ratio 1 schedule). One of the levers then again became available after a delay composed of a constant duration plus a variable one. The constant component was 5 s, and the variable one was sampled from a truncated Poisson distribution with a mean of 20 s and a

range of 5 – 60 s. All three sessions concluded after 60 reward deliveries i.e., 30 lever presses on each side, or after 3 h.

Cue training

To train the subjects to the reward contingencies of the four auditory cues they would encounter in the main experiment (see *Treatment Groups* and *Choice Alternative* sections below), we conducted a Pavlovian protocol in which all the rats were exposed to the cues and their respective reward contingencies. In this phase, cue presentation was independent of the behaviour of the rat. These cue-training sessions consisted of 40 trials, with 10 trials for each of the 4 cues, intermixed in random order. To avoid large deviations from the expected outcome probabilities of cues N1 and N2 in each session, proportions of outcomes were fixed as one half for each cue, meaning in each session each cue was followed by 3 reward pellets on 5 trials and 1 reward pellet on 5 trials. Trials were separated by an ITI generated by sampling from of a truncated Poisson distribution with a mean of 50 s (range 10–120 s) + 10 s (to ensure a minimum ITI of 20 s; range: 20 s $\leq$ ITI $\leq$ 130 s). Subjects performed one daily session of this phase for 15 days. Number of head-pokes into the food magazine during cue presentation was measured to establish the degree of cue discrimination.

Experimental procedures

General procedure

We used a trial-based chain procedure as displayed in **Fig. 1**. There were two kinds of trials: 2-option choice trials and 1-option forced trials. A day's session was composed of 60 trials: 40 forced and 20 choice, which were randomly intermixed. Subjects performed one daily session for 14 consecutive days. Each treatment group had a unique combination of two out of three possible choice alternatives (see *Treatment Groups* and *Choice Alternative* sections below). All trials started with the rear lever (initial link) extending. Pressing this lever resulted in its retraction, and either one (forced trials) or both (choice trials) of the front levers being presented. Pressing a front lever initiated an acoustic cue (the terminal link) and the retraction of that lever (forced trials) or both levers (choice trials). The auditory cue was broadcast for a 10 second interval, after which food delivery occurred (3 sucrose pellets, 2 sugar pellets or 1 sucrose pellet, depending on the choice alternative). Trials were separated by an inter-trial interval (ITI) generated by sampling from of a truncated Poisson distribution with a mean of 50 seconds (range: 10 – 120 seconds) and adding 10 seconds. A session finished after 60 trials or 2 hours, whichever occurred first. Each rat was trained at the same time every day; one cohort of rats began the experiment at 9:00 A.M., another at 12:30 P.M., and the last at 3:30 P.M.

Treatment Groups

A between-subjects design was used, with eight rats in each treatment group. Subject assignment to group was organised such that there was no correlation between group and any of the following parameters: side of a choice alternative; hour of testing; cage in which the animals were housed, or cue-reward contingencies. Subjects in each treatment chose between a unique combination of two alternatives, reflected in the group name.

In *Risky(Info) vs Risky(NoInfo)*, subjects were presented with two options that were risky i.e., the delayed reward outcome of a given trial (3 or 1 sucrose pellet(s)) was stochastic and could not be determined by subjects pre-choice, and sequences of reward outcomes across trials could not be predicted ahead of time. In Risky(Info), the forthcoming reward outcome of a given trial was signalled immediately post-choice by auditory cues, while in Risky(NoInfo), the forthcoming reward outcome was uncertain until the 10s delay time preceding the reward outcomes lapsed. Thus, while both options involve risk in that outcomes are unpredictable pre-choice, they differ in that Risky(Info) provides post-choice certainty within a given trial while Risky(NoInfo) does not.

In *Risky(NoInfo) vs Fixed* subjects chose between Risky(NoInfo) and a Fixed option which always delivered 2 sucrose pellets 10s after being selected. There

was no risk associated with the Fixed option since the reward outcome could be predicted pre-choice.

In *Risky(Info) vs Fixed* subjects chose between the Risky(Info) option and the Fixed option.

Choice Alternatives

Overall, there were 3 choice alternatives, but each subject was exposed to 2 of these alternatives, the combination of which depended on the treatment group. All choice alternatives had the same average profitability (expected food/ expected time).

Selection of Risky(Info) resulted in an auditory stimulus, either S+ or S-, being broadcast for 10s with equal probability, followed by either 3 sucrose pellets or 1 sucrose pellet respectively. S+ was always followed by 3 pellets and S- was always followed by 1 pellet, hence subjects were informed of the magnitude of trial outcomes immediately post-choice (note that in conditioning terms S- generally refers to a stimulus that signals no reward, whereas here we define S- as the stimulus signalling the least reward compared with others, but not zero. We adopt the label here to maintain consistency with the previous literature such that the stimulus signalling the least reward is designated S-).

Selection of Risky(NoInfo) resulted in an auditory stimulus, either N1 or N2, being broadcast for 10s with equal probability. Each stimulus had a 50% chance of resulting in either 3 sucrose pellets or 1 sucrose pellet, and therefore did not provide advance information about the magnitude of forthcoming reward outcomes.

Selection of Fixed resulted in an auditory stimulus, either S1 or S2, being broadcast for 10s with equal probability. Each stimulus was always followed by 2 sucrose pellets. Because there was no pre-choice uncertainty associated with the Fixed option S1 and S2 did not provide additional information regarding the magnitude of reward outcomes.

Data Analysis

Data processing and analyses were carried out in MATLAB 2019a, and statistical tests were carried out with R statistical software (Version 4.2.3). In the analysis of response latency data in 1-option forced trials the median latency of each alternative was calculated for each subject across a session and used for subsequent analysis. A type-1 error rate of 0.05 was adopted for all statistical comparisons and a Bonferroni correction was applied for all multiple pair-wise comparisons. For statistical analyses, head-poking data were log transformed and

choice proportion data were arc-sine square-root transformed to normalize the residuals (Grafen and Hails, 2002).

Ethical Note

All experiments were carried out in compliance with the UK Animal (Scientific Procedures) Act (1986) and its associated guidelines.

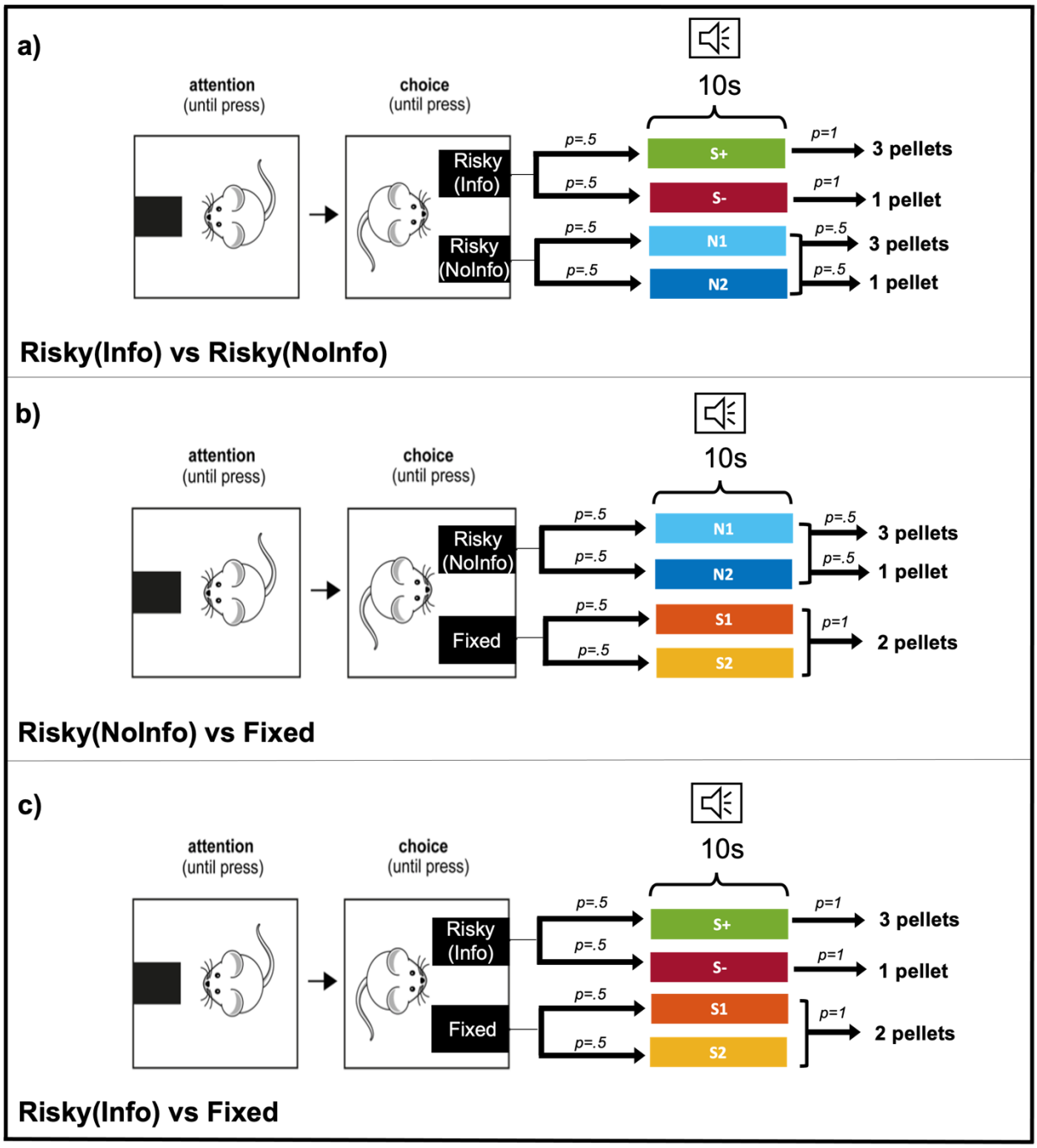


Fig. 1 Experimental design showing choice trial structure for each group. Black arrows show transition points within a trial. A trial begins with the presentation of the back lever which when pressed retracts and generates choice levers (one or both). Choice is followed by the presentation of an auditory cue for 10s, and then the trial reward outcome. *p* denotes probability.

Results

Cue Training

Fig. 2 shows the average number of head pokes made during the 10s interval preceding reward outcomes when the cues were present, with data pooled from the last three training sessions for each group.

In *Risky(Info) vs Risky(NoInfo)*, the number of head pokes ranked according to the average magnitude of reward associated with each cue (**Fig. 2a**). The number of head pokes was greatest when S+ was presented (37.0 ± 2.9; mean ± s.e.m.), intermediate when N1 and N2 were presented (33.5 ± 2.8, 32.6 ± 2.9, respectively) and lowest when S- was presented (28.4 ± 2.1). A one-way repeated-measures ANOVA revealed a significant effect of cue ($F_{3,85} = 9.49$, $P < 0.0001$). Post-hoc pair-wise comparisons showed a significant difference in head poking between S+ and S- ($t_7 = 5.99$ $P < 0.01$), but not between either of these cues and N1 and N2, or N1 and N2 themselves (which is expected given that these two cues are associated with the same average reward magnitude).

(**Fig. 2b**) shows that in *Risky(NoInfo) vs Risky(Fixed)*, the number of headpokes was similar for all cues (N1: 38.5 ± 2.6, N2: 36.0± 2.9, S1: 37.3± 2.6, S2: 37.3±

2.8), which is as expected given that they signal the same average reward magnitude. This was confirmed by a one-way repeated-measures ANOVA that showed no significant effect of cue ($F_{3,85} = 1.47$, $P = 0.23$).

In *Risky(Info) vs Fixed*, the number of head pokes ranked according to the average magnitude of reward associated with each cue (**Fig. 2c**). The number of head pokes was greatest when S+ was presented (41.5 ± 2.3; mean), intermediate when S1 and S2 were presented (39.6 ± 2.4, 38.5 ± 2.7, respectively) and lowest when S- was presented (31.9 ± 1.7). A one-way repeated-measures ANOVA revealed a significant effect of cue ($F_{3,85} = 18.40$, $P < 0.0001$). Post-hoc pair-wise comparisons showed a significant difference in head poking between S+ and S- ($t_7 = 9.12$ $P < 0.0001$), S- and S1 ($t_7 = -6.21$ $P < 0.01$) and S- and S2 ($t_7 = -4.34$ $P < 0.05$), but not between S+ and either S1 and S2, or S1 and S2 themselves (the latter result being expected given that these two cues are associated with the same average reward magnitude).

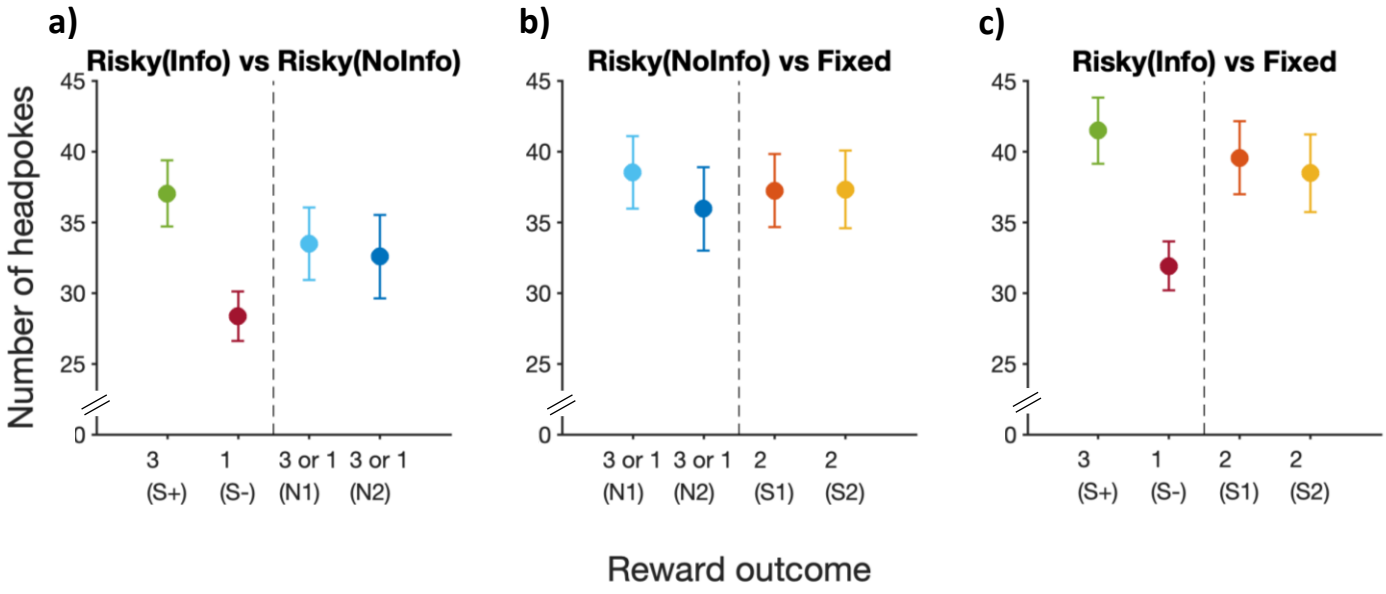


Fig. 2 Number of head pokes into the food magazine during cue presentation at the end of the training phase. The data shows the average number of head pokes (mean ± s.e.m.) made by subjects over the 10s intervals preceding reward outcomes, pooled from the last three sessions of training. During this time reward-predictive signals indicating the magnitude of forthcoming reward outcomes were presented. **a)** Shows data from *Risky(Info) Vs Risky(NoInfo)* **b)** Shows data from *Risky(NoInfo) Vs Fixed* and **c)** shows data from *Risky(Info) Vs Fixed*. n=8 per group.

Experiment

Preference 1: choice in 2- option trials

Fig. 3 shows that across all three treatments, on average, subjects developed a preference for one of the two alternatives available to them, over the 14 sessions of the experiment, albeit with differing strengths. A two-way repeated-measures ANOVA with choice proportion as the dependent variable and group and session as independent variables revealed a significant effect of group ($F_{2,21} = 19.47$, $P < 0.0001$), session ($F_{13,273} = 4.96$, $P < 0.0001$), and a significant interaction ($F_{26,273} = 5.61$, $P < 0.0001$).

In *Risky(Info) vs Risky(NoInfo)*, subjects developed a very strong preference for Risky(Info) **(Fig. 3a**). Pooled over the last 3 sessions group preference for Risky(Info) reached 0.99± 0.0034; mean ± s.e.m. In *Risky(NoInfo) vs Fixed*, as a group subjects displayed a mild preference for the Fixed option **(Fig. 3b**), reflected in a low preference for NoInfo over the last 3 sessions (0.28± 0.073). In *Risky(Info) vs Fixed*, subjects developed a very strong preference for Risky(Info) **(Fig 3c**), with preference for Risky(Info) reaching 0.98 ± 0.0072 over the last 3 sessions. Taken together these results reveal the following average ranking of choice alternatives by our subjects: Risky(Info) > Fixed > Risky(NoInfo).

Preference 2: latency in 1- option trials

Fig. 3 also shows how quickly subjects responded to each available alternative when they were presented alone in forced trials. Latencies, or reaction times in 1-option trials have proven to be a robust predictor of choice in 2-option trials, and are very informative with respect to the psychological mechanism of choice (see, for instance, Monteiro et al., 2020), offering an independent measure of preference. Shorter response times are interpreted to reflect a greater preference.

Fig. 3 shows that latencies in single-option trials mirrored the rats' choice proportions in two-option choice trials. During the last 3 sessions of the experiment, the option chosen most frequently had the shortest response latency. This was the case in *Risky(Info) vs Risky(NoInfo)* (**Fig. 3d**; 2.00 ± 0.093 vs 2.69 ± 0.12; t_7 = -4.64, $P < 0.01$), *Risky(NoInfo) vs Fixed)* (**Fig. 3e**; 3.51 ± 0.26 vs 3.10 ± 0.18; t_7 = 1.58, $P = 0.159$) and *Risky(Info) vs Fixed* (**Fig. 3f**; 1.90 ± 0.097 vs 2.72 ± 0.12; t_7 = -5.05, $P < 0.01$).

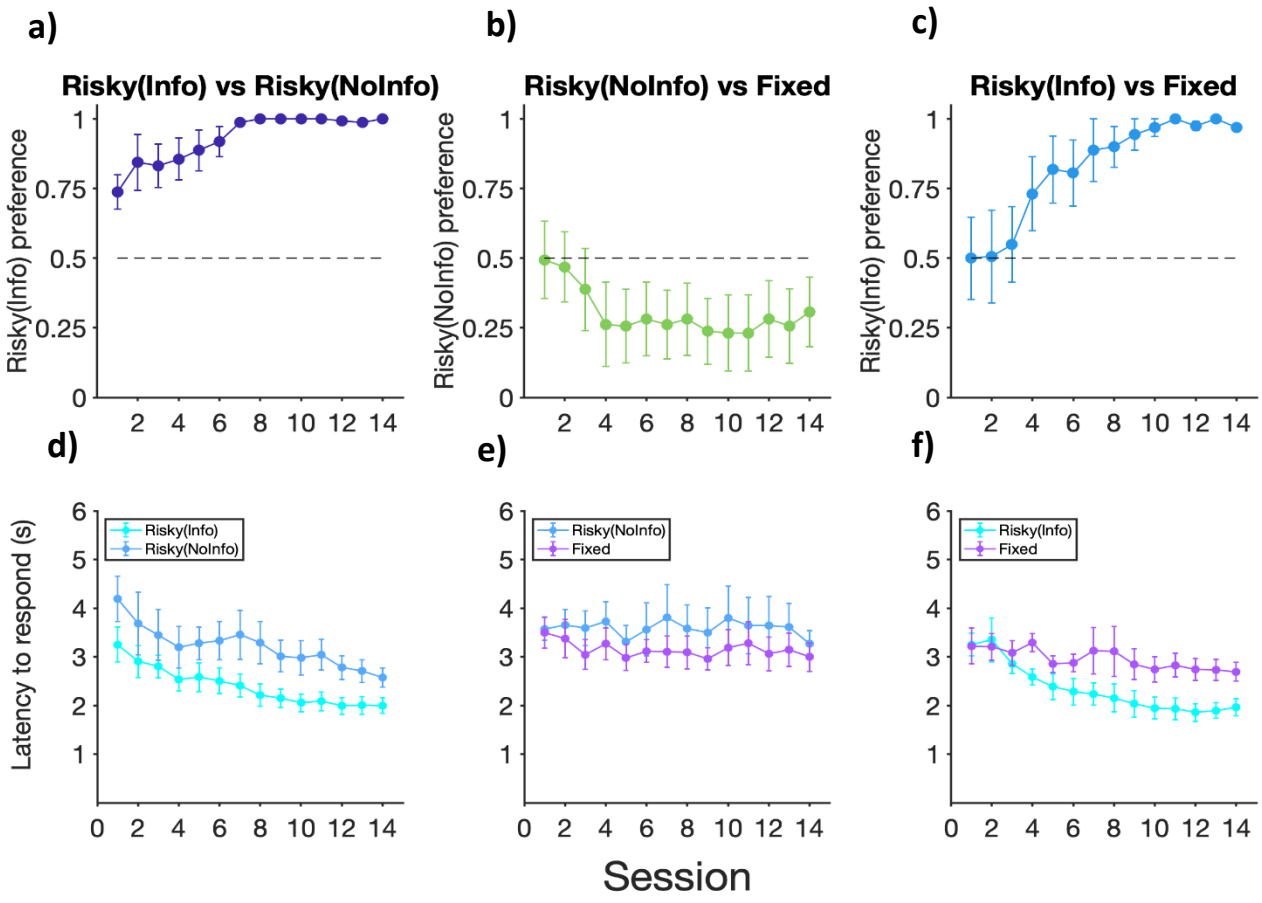


Fig. 3 Option preferences as choice proportions in 2-option choice trials and response latencies in 1-option forced trials. a) Choice proportions for the Risky(Info) option when presented with Risky(NoInfo). **b)** Choice proportions for the Risky(NoInfo) option when presented with Fixed. **c)** Choice proportions for the Risky(Info) option when presented with Fixed. **d)** Response latencies for *Risky(Info) Vs Risky(NoInfo)*. **e)** Response latencies for *Risky(Info) Vs Fixed*. **f)** Response latencies for *Risky(Info) Vs Fixed*. means ± s.e.m. n=8 per group.

Discussion

The aim of this study is to examine how risk in amount of reward (i.e., quantifiable unpredictability), and the availability of advanced information delivered by reward-predictive signals, interact to generate preferences. This was motivated by two well-documented and, at times, incongruous empirical observations from historically distinct experimental paradigms. The first is that animals tend to be risk averse for amounts of reward (Kacelnik and Bateson, 1996), and the second is that individuals show robust preferences for risky alternatives presenting reward-predictive stimuli that deliver advanced information about upcoming stochastic rewards (reviewed in Dunn et al., 2023). This raises the broader question of how individuals trade-off preferences pertaining to stochastic variability, with the drive to seek informative signals, in order to maximise fitness.

Across three treatments subjects chose between two alternatives of equal average profitability (reward/time) that only differed in the availability of advanced information about stochastic reward magnitudes; *Risky(Info) vs Risky(NoInfo),* variance in reward magnitude; *Risky(NoInfo) vs Fixed,* or both; *Risky(Info) vs Fixed.* We recorded two metrics of preference: choice proportions in 2-option trials when the alternatives were presented simultaneously, and latency to

respond in 1-option trials when subjects were forced to take a single option. The latter kind of trial serves to expose individuals to the contingencies of each option, eliminating the possibility that side biases preclude learning about the relevant alternatives. Latencies to respond in forced trials also provide an independent metric of preference that has proven to be more sensitive than choice proportions (e.g., Brito e Abreu and Kacelnik, 1999; Reboreda and Kacelnik, 1991) and is also informative regarding general mechanisms underlying choice (see Ajuwon et al., 2022; Macias et al., 2020; Monteiro et al., 2020; Ojeda et al., 2018). Shorter response latencies are interpreted to reflect a greater degree of preference.

Consistently with previous studies on paradoxical choice in *Risky(Info) vs Risky(NoInfo)* rats showed a robust preference for the informative option offering reward-predictive signals about forthcoming stochastic outcomes (**Fig. 3a,d**). Moreover, in *Risky(NoInfo) vs Fixed* our results are consistent with the empirical trend that a range of animals are weakly risk averse for amounts of reward, preferring a fixed option over a variable one. We found that most, but not all of our subjects preferred the fixed option (**Fig. 3b,e**). Most strikingly, when both of these factors differed between available alternatives, subjects showed a strong preference for the risky, but informative option (**Fig. 3c,f**), as if the observed preference for informative signals in *Risky(Info) vs Risky(NoInfo)*, supplanted the risk aversion for amounts of reward displayed in *Risky(NoInfo) vs Fixed.* It is unlikely that these results can be accounted for by a putative systematic inability

of our subjects to discern the contingencies of the options available to them. During the last 3 sessions of the training phase when subjects were exposed to the reward predictive stimuli of each option, their head-poking responses accurately reflected the average pay-off associated with each stimulus **(Fig. 2)**.

Across the three groups our subjects showed the following transitive (internally consistent; see Monteiro et al., 2013) ranking of choice alternatives: Risky(Info) > Fixed > Risky(NoInfo), highlighting a strong preference for Risky(Info) - the only option providing unambiguous informative signals regarding the magnitude of reward outcomes. While a number of studies have shown that rats show preferences for informative signals anticipating the presence or absence of reward (Ajuwon et al. 2022; Cunningham & Shahan 2019; Ojeda et al., 2018), here we show that these preferences extend to information pertaining to the magnitude of reward, a result that has also been found in macaques (Bromberg-Martin and Hikosaka, 2009, 2011; Blanchard et al., 2015). This indicates that the mechanisms motivating the display of informative signals are not exclusive to one dimension of reward-related uncertainty. Future work should aim to better characterise the kinds of advanced non-instrumental information that subjects are sensitive to, which may well extend beyond foraging contexts.

Several mechanisms may account for robust preferences for the informative option, however our results do not support the hypothesis that an aversion to temporally prolonged uncertainty can promote these preferences, as proposed in a computational cognitive model by Bennett et al. (2016). They argued that information derives primary reinforcing value because it provides a relief from reward related uncertainty that is extended through time. However, this predicts that in *Risky(Info) vs Fixed* subjects should avoid the pre-choice uncertainty associated with the variable alternative altogether, and exclusively prefer the fixed option which is associated with no uncertainty. In terms of information acting as a primary reinforcer, our results are better accounted for by mechanisms emphasising that the resolution of uncertainty is appetitive (Bromberg-Martin and Hikosaka, 2009; Daddaoua et al., 2016; Hendry, 1969). The distinction is subtle but we think that it is significant: mechanisms involving an appetitive information drive could lead individuals to actively interrogate natural environments in search of specific kinds of information, while an aversion to uncertainty may have the opposite effect in natural settings, keeping individuals restricted to fully predictable environments in which no learning occurs. This makes accounts proposing an appetitive influence of uncertainty resolution appealing owing to greater ecological realism.

Bennett et al.'s (2016) model does however, emphasise another factor that has proven an important modulator of preferences for advanced information: time under uncertainty. Vasconcelos et al. (2015) found that when reward-predictive stimuli in the informative option were presented just before outcomes at the end of the delay period rather than immediately after choice, preference for the informative option vanished. This indicated that while the presence of reward predictive stimuli are important, the timing of their delivery matters also. One way to account for this is to consider that the resolution of uncertainty is appetitive, but that like other primary reinforcers, its value may be discounted over time (Kacelnik, 2007, 2003) so that the *early* resolution of uncertainty is more desirable. Future work should seek to formalise this proposal, which could be investigated using experimental techniques that have been developed to study the discounting of future food rewards in animals.

While there are intuitive functional reasons to subscribe to the hypothesis that the early resolution of uncertainty is appetitive, purely information-based accounts of paradoxical choice fail to capture an important empirical observation - that signals for good news appear to have a greater influence on preference than signals for bad news (Ajuwon et al., 2022; Fortes et al., 2017; Laude et al., 2014; McDevitt et al., 1997; Pisklak et al., 2015; Spetch et al., 1990, 1994; for reviews see Cunningham and Shahan, 2018; Dunn et al., 2023; Zentall 2016, and Dinsmoor 1983 for a similar finding in the observing response literature). This is

consistent with other studies showing that subjects attend to Pavlovian cues for 'good news' in contexts where they are not rewarded (Castro and Berridge, 2014). An alternative hypothesis that is consistent with our results and the asymmetric effect of stimuli of opposing valence in other studies, is the idea that individuals find the anticipation of desirable outcomes hedonically attractive, and that reward prediction errors associated with these outcomes boost anticipation (Ligaya et al., 2016, 2020). This hypothesis, formulated as a reinforcement learning model, also accounts for an otherwise perplexing finding whereby longer times between the choice response and variable outcome increase preference for the informative option (Cunningham and Shahan 2019; Ligaya et al. 2016, 2020; McDevitt et al., 2018).

In *Risky(NoInfo) vs Fixed*, when alternatives differed only in the variance in amounts of reward and not in advanced information, we found that generally subjects were risk averse, consistently with preferences in a range of species (Kacelnik and Bateson, 1996; Kacelnik & El Mouden, 2013; Craft , 2016). This result confirms the predictions of Scalar Utility Theory (SUT), a process-based account of risk sensitive preferences in animals, arguing that risk sensitivity is a by-product of information-processing mechanisms that have not evolved in response to risk *per se,* but as an adaptive compromise allowing individuals to perceive and respond to a range of stimuli across a broad magnitude range. A corollary of having the ability to sense stimuli over a range of several magnitudes

is that the difference required to distinguish stimuli of greater intensity (referred to as the just noticeable difference) increases with absolute magnitude of the stimuli (Foley and Bates, 2019). Another way of thinking of this is that stimuli of greater intensity are perceived and remembered less precisely than stimuli of smaller intensity. SUT therefore assumes that for reward amounts, this error is proportional to stimulus magnitude. When an option varying in reward amounts is represented in memory, this relation between error and magnitude leads a variable option to be remembered by an animal as having a smaller median (but still same mean) reward than a fixed option objectively giving the same average reward as the variable one (Kacelnik and Brito e Abreu, 1998).

Preference for the variable option in *Risky(NoInfo) vs Fixed* can also be accounted for by mechanisms of associative learning. If the reinforcement value of rewards is a concave function of reward amount, then Jensen's inequality implies that variance in amount will lead subjects to prefer a fixed option over a variable one of equal average reward rate (Kacelnik and Bateson, 1997). Interestingly, a similar argument based on secondary reinforcement has been made to explain preference for the informative option in paradoxical choice (our *Risky(Info) vs Risky(NoInfo)* condition) and related observing response experiments (Dinsmoor, 1983; Roper and Zentall, 1999). In this case, however, it is suggested that the reinforcing value of a stimulus predicting reward is a convex function of the reward magnitude, meaning that subjects prefer the alternative with stimuli

reliably predicting a large and small (or no) reward vs one with an ambiguous stimulus associated with the average reward amount. Clearly, these two mechanistic accounts of behaviour, originally developed to explain different empirical observations, are incompatible: it is rather unparsimonious to conclude in our experiment that while the function relating primary reinforcement value to reward magnitude is concave, the function relating secondary reinforcement value to signals for reward is convex, given that the signals are assumed to derive their reinforcing value from the reward they are associated with this. Based on the present results, we therefore surmise that the robust preference for Risky(Info) is most likely caused by an appetitive drive to resolve uncertainty, or the boosting of anticipation by reward prediction errors for desirable outcomes, rather than conditioned reinforcement.

In *Risky(Info) vs Fixed* we set out to test the interaction between the effect of variance in amount of reward and the availability of informative stimuli, on choice. We found that though subjects showed risk aversion in the classic risk protocol, when stimuli anticipated the magnitude of forthcoming rewards in the risky option immediately after the choice response, subjects showed a robust preference for the risky option. This is consistent with previous findings that reward predictive stimuli can increase preference for a risky over a fixed option (Case and Zentall, 2018; Zentall and Stagner, 2011) and that they can more generally promote risk proneness (Barrus and Winstanley, 2016). The switch

from risk aversion to risk proneness that we found between cannot be explained by current normative or mechanistic accounts of risk sensitivity. As previously discussed, Scalar Utility Theory predicts risk aversion for amounts of reward and not risk proneness (but see Rosenström et al., 2016 for a dissenting view). Classical Risk Sensitivity Theory does predict shifts between risk aversion and proneness, but only in response to a decrease in subjects' energy budget, and all of our subjects were kept under the same feeding regime. The switch in preference to risk proneness can be attributed to a putative reinforcing effect of the resolution of uncertainty or a boost in anticipatory utility afforded by reward predictive signals. These results highlight that theoretical accounts of risk sensitivity should include effects of advanced information about variable outcomes, and be expanded to consider the wider trade-offs that subjects face when making risky decisions. Specifically, in natural settings, individuals may show a greater degree of risk proneness for amounts of reward than previously predicted, owing to post-choice uncertainty reduction via reward predictive signals in the environment, which may reduce the perceived and experienced riskiness of alternatives. Also, departures from risk aversion may reflect a broader trade-off between optimisation under risky choice and a drive to acquire information. These results highlight that the availability of advanced information and its value can modulate pre-existing biases in a range of contexts, though the functional implications are not entirely clear yet.

Supplementary materials

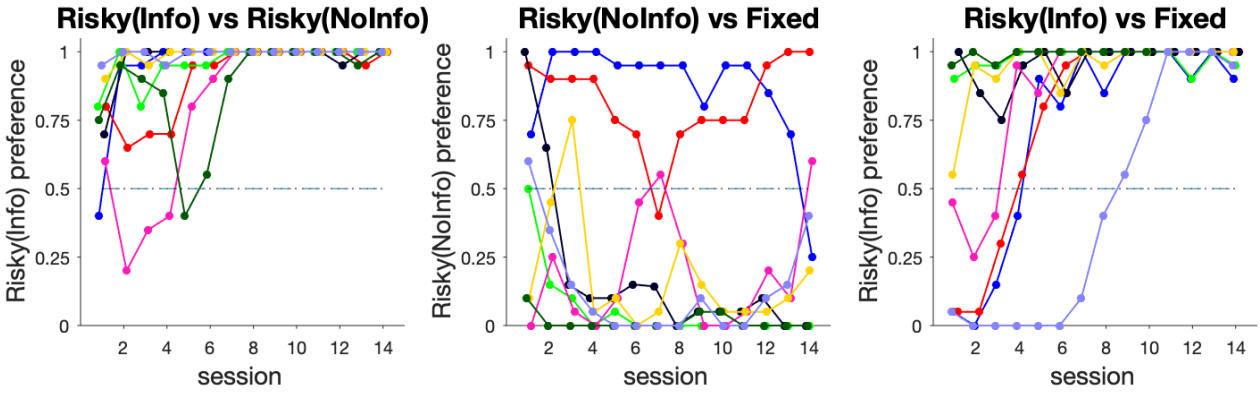


Figure S1. Individual preferences for the risky over fixed alternative. Proportion of choice are shown with n=8 in each group. Each colour represents a different individual for each group.

Author List: Victor Ajuwon, Andres Ojeda, Tiago Monteiro, Mark Walton, and Alex Kacelnik

Author contributions: VA, AO, MW and AK conceptualised and designed the experiment; VA collected the data; VA, TM, MW and AK analysed the data; VA wrote the first draft of the paper; all authors are currently reviewing the manuscript.

Chapter 3. *GoFish*: A low-cost, open-source platform for closed-loop behavioural experiments on fish

A manuscript published in Behavior Research Methods

Abstract

Fish are the most species-rich vertebrate group, displaying vast ecological, anatomical and behavioural diversity, and therefore are of major interest for the study of behaviour and its evolution. However, with respect to other vertebrates, fish are relatively underrepresented in psychological and cognitive research. A greater availability of easily accessible, flexible, open-source experimental platforms that facilitate the automation of task control and data acquisition may help to reduce this bias and improve the scalability and refinement of behavioural experiments in a range of different fish species. Here we present *GoFish,* a fully-automated platform for behavioural experiments in aquatic species. *GoFish* includes real-time video tracking of subjects, presentation of stimuli in a computer screen, an automatic feeder device, and closed-loop control of task contingencies and data acquisition. The design and software components of the platform are freely available, while the hardware is open-source and relatively inexpensive. The control software, *Bonsai*, is designed to facilitate rapid development of task workflows and is supported by a growing community of users. As an illustration and test of its use, we present the results of 2 experiments on discrimination learning, reversal, and choice in goldfish (*Carassius auratus)*. *GoFish* facilitates the automation of high-throughput protocols and the acquisition of rich behavioural data. Our platform has the potential to become a

widely used tool that facilitates complex behavioural experiments in aquatic species.

Introduction

A common framework for the study of animal behaviour and cognition involves presenting stimuli and manipulanda, measuring animals' movements and responses, and programming outcomes according to selected contingencies, while recording all of this information, including fine temporal details. These efforts have led to the development and use of conventional experimental platforms that satisfy these needs while ensuring replicability across laboratories. Perhaps the archetype of such a system is the Skinner box, originally designed for pigeons and rodents, that uses manipulanda suitable for those taxa, detecting behaviour by the closing of circuits through key-pecking, lever-pressing, or interruption of light beams. Such systems promote and enhance reproducibility, a critical need in contemporary behavioural research. However, studying the behaviour of organisms living under water, such as fish, cephalopods, and crustaceans, poses different technical challenges from those in terrestrial species.

Unlike most behavioural laboratory-based experiments involving mammals and birds, the display of stimuli and delivery of food reinforcers for fish is frequently manually executed by an experimenter, increasing temporal variability and vulnerability to observer effects, while restricting scalability (e.g., Potrich et al., 2022; Schluessel et al., 2022). Similarly, data are often recorded by video but

annotated visually or digitised at a later time instead of being processed in real-time which allows behaviour to control reward through pre-programmed contingencies.

In spite of their huge ecological, neuroanatomical and behavioural diversity, historically fish have been underrepresented in the cognitive and psychological literatures for decades (Bitterman, 2006; Newport, 2021; Shettleworth, 2009), with primates, corvids, pigeons, rodents and more recently dogs being favoured models (though see Bitterman, 1975; Chase & Hill, 1999; Gerlai, 2014; Pouca & Brown, 2017; Salena et al., 2021). Increasing the availability of user-friendly, open-source, experimental platforms that allow for automated testing and data acquisition in fish may help to mitigate this bias (Brock et al., 2017; Gatto et al., 2021).

Advances in zebrafish research highlight how novel technology, increased experimental automation (Aoki et al., 2015; Brock et al., 2017; Guilbeault et al., 2021; Kuroda et al., 2017; Manabe et al., 2013; Miletto Petrazzini et al., 2020; Mueller & Neuhauss, 2012; Santacà et al., 2021; Stewart et al., 2015), and an ideal model fish species, together can help to unravel the links between genes, neural activity, behaviour and cognition (for reviews see Gerlai, 2014, 2020; Kalueff et al., 2013; Meshalkina et al., 2017; Orger & de Polavieja, 2017), with translational impacts for the treatment of disease (Kalueff et al., 2014).

Nonetheless, despite the many advantages of zebrafish as models (e.g., their rapid life-cycle, optical transparency enabling non-invasive whole-brain electrophysiological recordings, and a variety of sophisticated genetic tools) the use of zebrafish for more cognitively demanding tasks may be limited (Blaser & Vira, 2014; but see: Gerlai, 2017; Santacà et al., 2021; Sridhar et al., 2021).

Recent attempts (i.e., <10 years) to improve the automation of behavioural experiments in other fish species using closed-loop systems have shown promising results. For example, Wallace et al (2020) investigated sex differences in numerical discrimination abilities in mosquitofish using an automated setup that facilitated a range of cognitive tests. Furthermore, automated systems that were originally developed for conditioning experiments in zebrafish (Gatto, Lucon-Xiccato, et al., 2020; Kuroda et al., 2017; Manabe et al., 2013) have been co-opted for use in guppies (Gatto et al., 2021; Gatto, Testolin, et al., 2020; Lucon-Xiccato et al., 2018).

However, most of these systems are either commercial solutions and therefore not easily adaptable or accessible (owing to higher costs), or are open-source but require a considerable degree of expertise to operate and adapt, thus lacking the flexibility to be easily applied to multiple experimental situations and/or other subject species.

To address this, we developed *GoFish*, an open-source and expandable platform for dynamic, fully-automated behavioural experiments on fish or other aquatic organisms. Our aim with *GoFish* is to provide a platform facilitating high-throughput and highly reproducible research that is i) open-source, ii) relatively inexpensive and simple to assemble, iii) readily modifiable, iv) supported by a growing community of users, and v) capable of providing a range of behavioural metrics.

Our platform is inspired by present day behavioural, cognitive and neuroscience experiments that rely on open-source, community based, *DIY*-type solutions for running and developing new experimental paradigms, as well as for processing and analysing the resulting data streams (Akam et al., 2022; Aoki et al., 2015; Bishop et al., 2022; Buscher et al., 2020; Devarakonda et al., 2016; Geissmann et al., 2017; Guilbeault et al., 2021; Gurley, 2019; Kane et al., 2020; Kapanaiah et al., 2021; Lopes et al., 2021; Mathis et al., 2018; Oh et al., 2017; O'Leary et al., 2018; Pineño, 2014; Siegle et al., 2017; Štih et al., 2019; Swanson et al., 2021; Walter & Couzin, 2021).

Briefly, our system allows for the display of stimuli on a computer screen placed outside but adjacent to a tank, the tracking and detection of the subject's location in real-time through an overhanging camera, the programming of contingencies between fish movements and the delivery of food rewards, and the automatic

recording of data in analysable format. Here, we describe the system and present two closed-loop experiments aimed at demonstrating its performance as a research tool. Although we describe an implementation for goldfish, *GoFish* can, in principle, be used with other aquatic species with minimal modifications.

As a proof-of-concept, and inspired by classical experiments (Bitterman, 1975; Engelhardt et al., 1973), we ran two closed-loop discrimination experiments using real-time video tracking. We show that individual goldfish can be trained to (i) associate a signalled location with food reward and reverse preference appropriately when the contingencies are reversed (*Experiment* 1), and (ii) discriminate coloured visual stimuli that switch location between trials (*Experiment 2*).

Methods

The *GoFish* platform

The setup as presently implemented (Figure 1) comprises a rectangular prismatic experimental tank (60 x 30 x 36 cm (length x width x height), Table 1) with a 17'' LCD computer screen (1920 x 1080; 60 Hz) for stimulus presentation (Table 1), placed directly adjacent to the side of the tank where reward pellets are delivered, (Figure 1a). Two custom-made, automated pellet dispensers (i.e., feeders (Figure 1b,c, Table 1) are clamped onto the upper edge of the tank such that pellets fall on the water surface approximately 2 cm from the closest side of the tank, adjacent to the screen. Each feeder is placed on either side of an opaque, white acrylic divider (fixed with silicone sealant), running perpendicular to the LCD computer screen, 25 cm into the tank. This partition defines the 2 choice zones of a Y-maze configuration. An overhanging USB camera (1280x720 resolution, Table 1) held above the tank records each session (Figure 1a). A laptop (Table 1) controls task contingencies (stimulus presentation and reward delivery) and video acquisition with a Bonsai (Lopes et al., 2015, 2021; Lopes & Monteiro, 2021) custom workflow (see Figure 1d for an example workflow). A light source (Table 1) is placed outside the tank, opposite to the LCD computer screen (Figure 1a). The tank is surrounded by opaque styrofoam panels to visually isolate the

fish during experiments. The water level is maintained at approximately 15 cm. In the experiments described below, 2 identical experimental tanks were run concurrently, with each fish being tested always in the same tank.

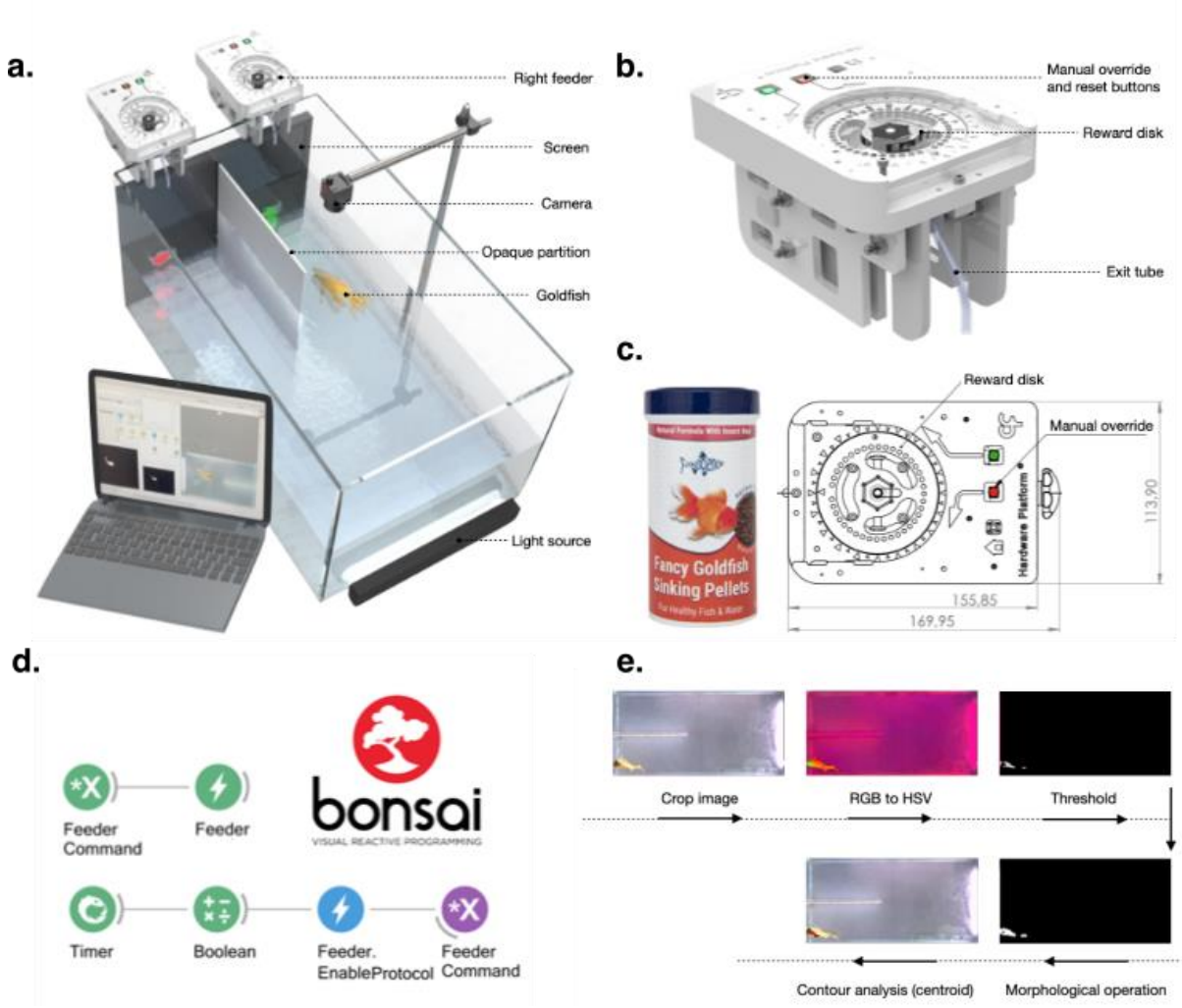


Figure 1. *GoFish* apparatus, pellet dispenser control and specifications, and video tracking pipeline. a. 3D view of closed-loop operant chamber. Setup includes two custom-made pellet dispensers, computer screen, usb camera and light source. **b.** 3D depiction of the pellet dispenser. **c.** Food rewards (left) and detailed top view of feeder reward container disk. **d.** Example *Bonsai* workflow for custom pellet dispenser control. The code implements periodic delivery of food pellets. **e.** Real-time video analysis pipeline.

Table 1. Parts list. Details of the components used to build one closed-loop behavioural chamber for goldfish learning experiments. Many of the components can be swapped for items of similar functionality to suit particular needs. Pellet dispenser parts list and assembly instructions can be found in a dedicated repository (see **Pellet dispenser** section for details). Prices are from early 2021, rounded to the nearest pound.

Component Name	Supplier	Reference code	Quantity	Unit price	Total Price
Clearseal All Glass Aquarium (24x15x12in; LxHxW)	Goldfish Bowl, Oxford, UK	N/A	1	£55	£55
WIFIGDS 17 Inch Monitor	Amazon	B08J7JJD3G	1	£81	£81
1080P 2MP ELP Web Camera	ELP	ELP-USBFHD08S-UK	1	£72	£72
Pellet dispenser	Champalimaud Hardware Platform	FF_GF_SP_V1.1	2	£142	£284
IREENUO Aquarium Light (11 inch)	IREENUO	HDMU-ATL-18A-DE	1	£14	£14
Laptop computer	Dell	Latitude 5520	1	£537*	£537
5mm White Acrylic (300x200x5mm)	Cut My Plastic	N/A	1	£10	£10
Betta Clear Silicone 310ml	Goldfish Bowl, Oxford, UK	N/A	1	£13	£13
Rankie USB 2.0 Extension Cable	Amazon	B01KWOBK84	3	£5	£15
Combined Total					**£1,081**

*UK educational price (VAT exempt)

Bonsai

The implementation of behavioural tasks and resulting data acquisition is controlled with *Bonsai*. *Bonsai* is a high performance, open-source visual programming software, for which there is an active community of thousands of users (https://github.com/bonsai-rx/bonsai/discussions) and several papers describing its inner workings (e.g., Lopes et al., 2015; Lopes & Monteiro, 2021). *Bonsai* allows users to rapidly develop workflows that can simultaneously manipulate data from various asynchronous input streams (e.g., video, or arduino controlled pressure sensors), while controlling numerous output devices (e.g., pellet dispensers).

Users can find documentation, video tutorials, online support and other materials in its accompanying website (BonsaiRX, http://bonsai-rx.org/). Briefly, Bonsai workflows are constructed by connecting functions, or 'operators' that come in the form of nodes, together (Figure 1d). These functions are categorised hierarchically within the Bonsai Toolbox that appears in the Bonsai workflow editor. For example, 'Source' functions allow users to easily generate data streams from files or external devices, while 'Sink' functions allow users to save data or trigger external outputs (https://bonsai-rx.org/docs/articles/editor.html). These functions can be searched for directly using the Search textbox that appears on top of the Bonsai Toolbox, saving users the need to search through all of the

functions within the Toolbox manually. A full list of Bonsai functions and their accompanying descriptions can be found here (https://bonsai-rx.org/docs/api/Bonsai.html). To quickly acquaint themselves with the basics of Bonsai, users can access common example workflows online (https://bonsai-rx.org/docs/tutorials/acquisition.html), or import them directly into their workflow editor through the Bonsai Gallery, which can be accessed via Tools in the menu bar of the workflow editor (https://bonsai-rx.org/docs/articles/gallery.html). Example video tutorials on how to quickly implement common workflows for data processing and storage can also be found here: https://bonsai-rx.org/learn/.

Pellet dispensers

Design and assembly instructions for laser cut acrylic and 3D printed parts for the pellet dispensers are available from the public repository (https://bitbucket.org/fchampalimaud/device.pump.fishfeeder/). The instructions include PCB manufacturing plans and specifications, as well as downloadable firmware. The dispensers are controlled through Bonsai (see example workflow in Figure 1d).

Stimuli

The potential visual stimuli and their positions are only limited by the monitor employed and its chromatic properties and dimensions. For the experiments described here, the main stimuli were coloured circles (red, green, blue and white, 3.5 cm in diameter, Figure 2b) on a grey background, presented with centres positioned 5 cm from the bottom of the tank and 7 cm from each side wall (Figure 1a). All stimuli were programmed using custom Bonsai (Lopes et al., 2015; Lopes & Monteiro, 2021) and BonVision code (Lopes et al., 2021) allowing easy generation and manipulation of visual stimuli. Each fish had a randomly assigned unique pair of colour-reward contingencies (Figure 2b). We chose colours that have been physiologically (Neumeyer, 1984) and behaviourally (Zerbolio & Royalty, 1983) proven to be discernible by our experimental species (goldfish, Carassius auratus). In a pre-experimental, pre-training phase (see details below), we used a white noise rectangle (13.5 x 12 cm, Gaussian: mean=0, variance=10) presented on either the left or right arms of the tank, or in both simultaneously, to signal the imminent delivery of reward in early pre-training, or to signal that reward delivery was contingent on fish swimming to a specific location in later pre-training stages.

Behavioural task control

Task control was fully automated and implemented using a custom Bonsai workflow (https://github.com/PTMonteiro/GoFish_Ajuwon_etal_2022). Progress through trials was controlled using real-time video analysis of fish movement. After a variable inter-trial-interval (ITI), fish could advance a trial by swimming into the 'start zone'. In the main experiment this was a 10 x 10 cm area opposite the rewarded side of the tank that was equidistant to both choice arms (Figure 2a). Presence in the start zone after the ITI would trigger stimulus presentation, and a subsequent crossing into either the 'left choice zone' or 'right choice zone' (15 x 15 cm; Figure 2a) would trigger appropriate contingencies. Outside of these epochs (and locations) the fish position had no influence on the unfolding of the task. Note that the 'start zone', 'left choice zone' and 'right choice zone' were not delineated by physical boundary markings, but were defined as specified regions of interest (ROIs) on the video feed corresponding to fixed areas within the experimental tank. Users may wish to make the ROIs visually identifiable to the subjects, as this may influence speed of acquisition. Frames from these ROIs were converted to HSV colour space and a HSV range was applied so as to successfully detect fish. The pixels of the resulting binarized frames from each ROI were summed continuously. Fish entry into the zones was recorded when summed ROI pixels exceeded a set threshold (the value of which was adjusted to each subject prior to the onset of the experiment).

After a session was completed a timestamped event list was generated as a *.CSV file. The first two columns indicate the event name as a string and a timestamp while the third column contains a number which encodes the particular outcome of events that require extra disambiguating information.

Fish tracking

To track the fish centroid in real-time a colour thresholding method (Monteiro et al., 2021) was implemented using a custom Bonsai workflow. Video was recorded at approximately 33 fps. Frames were cropped down to include only the inside of the tank and converted to HSV colour space. An HSV threshold was applied to isolate the fish body from the overall white background given by the tank's bottom. Prior to the onset of the experiment, HSV value ranges were manually set for each fish so as to provide robust tracking in spite of individual differences in fish coloration. The resulting binarized region (pixels are either fish or no-fish) was smoothed and the coordinates of the animals' centroid were extracted.

For each session a CSV data tracking file is generated which contains the x and y coordinates of the fish centroid throughout the session in two respective columns. A third column records the luminance of a specified ROI; a central point

at the top of the LCD screen on which stimuli are presented to subjects. Recording the luminance of this region provides information about the epoch of the task (Figure 2c) and therefore allows users to associate fish position with particular epochs of the task, enabling behavioural analysis during trial epochs of interest. For example, a low luminance value indicates an ITI period, an intermediate luminance value indicates the epoch during which a new trial is available and high luminance value represents the post-choice epoch within a trial. Example occupancy data from a representative experimental session, can be found in Figure 2d.

At the session end a raw video file of the entire session is also generated, allowing users to perform further tracking analysis offline.

Subjects

Five goldfish ranging in size between 7 – 10 cm, (age and sex unknown) participated in the current study. Animals were obtained from a local, commercial supplier (Goldfish Bowl, Oxford, UK).

Fish were housed in groups of two or three, in holding aquaria (60 x 35 x 31 cm; (length x width x height)) where they had access to a rock shelter, pebbles and artificial plants. They participated in experiments 5 times a week on weekdays and fed a total of 24 sinking pellets a day (Fancy Goldfish Sinking Pellets, Figure 1c). This diet was supplemented with spinach following experiments on the last day of the week and bloodworms the day after. Fish were kept under a 12:12 h light: dark cycle using fluorescent lights. Water was maintained at a minimum of 21°C using an internal heater and independent thermometer (pH: 8.2; ammonia: 0 ppm; nitrite: 0 ppm; nitrate: max. 30 ppm). Partial water changes were conducted at the end of each week and internal filters were cleaned every month. Each holding tank was aerated using an air pump.

For each daily session, fish were transported in a plastic jug to its experimental tank and then back to its holding tank at the end of the session. At the start of each day ~20L of water from all holding tanks were transferred to the experimental tanks in order to keep the environmental conditions as constant as

possible. The experimental tanks were cleaned at the end of each week. All animals had experimental experience with unrelated contingencies.

Pre-training

Pre-training consisted of three phases lasting a minimum of 18 days in total. Advancing through the phases depended on the individual subject's performance.

(i) Experimental tank acclimatisation

During a 10 min period, fish were allowed to explore and get acclimated to the tanks, previously baited with 12 food pellets throughout. This phase lasted for one day.

(ii) Choice zone training

The aim of this phase was for subjects to learn that swimming into either the left or right choice zone (outside of the ITI) was reinforced. After the ITI (drawn from a uniform distribution: min = 5 s; max = 10 s) during which the screen was black, a white noise rectangle would signal potential food availability in either the left or right choice zones of the tank. Reward was then contingent on fish entering the choice zone signalled with the white noise stimulus. For the first 5 days of

this phase there was one session of 12 trials per day and in the following five days, one session of 16 trials per day. Following this, for 3 days fish completed 2 sessions of 12 trials each per day. Rewards were evenly split across both choice zones and allocated randomly. A session ended either when all trials were completed or after 30 minutes.

(iii) Start position training

The aim of this phase was for subjects to learn that a start position had to be entered before subsequent behaviour could be reinforced. After the ITI (drawn from a uniform distribution: min = 20 s; max = 40 s), trial availability was signalled by a grey screen. During this period, fish were required to swim first to the back half of the experimental tank into a 'start zone' (i.e., >30 cm, away from the monitor and feeders) to trigger the onset of the white noise stimulus signalling food availability in either the left or right arm. As in the previous phase, reward was then contingent on fish entering a choice zone signalled with the white noise stimulus. This lasted for a minimum of 3 days. Following this, the start zone length was reduced in half (minimum 3 days), and finally to a 10 x 10 cm centred square (minimum 5 days) that was used in the main experiments (Figure 2a). There were 2 sessions of 12 trials each per day. To advance through this phase animals had to successfully consume the 12 food pellets within a 1h limit in each daily session. Failure to do so would terminate the training session, with fish

returned to their holding tanks. The remaining food pellets would be made available by the end of the day in the holding tanks.

Experiment 1: Acquisition and reversal of a spatial conditioning

Acquisition Phase

Every fish was presented with one daily session of 24 trials. A trial started with an ITI (drawn from a uniform distribution: min = 20 s; max = 40 s) where the screen was black and behaviour had no consequences. The ITI offset was signalled by a grey screen (Figure 2c) and from this moment on, entering the start position (Figure 2a,d) would trigger the presentation of both visual stimuli (i.e., S+ and S-, see Stimuli above) at fixed left/right locations (Figure 2b,c, counterbalanced across subjects). Fish made choices by entering one of the two choice zones (Figure 2a). Choosing the S+ side resulted in the delivery of a food pellet after a 5 s delay and the onset of an ITI. Conversely choosing the S- side would start a new ITI after a 5 s delay (Figure 2c). This experimental phase lasted for 5 days.

Reversal Phase

This phase followed the same contingencies as acquisition, except that the rewarded side for each animal (and accompanying stimuli location) was swapped, remaining the same after that. This phase lasted for 7 days.

Experiment 2: Colour discrimination

In this experiment the rewarded side (and S+/S- stimuli) was randomised on a trial-by-trial basis. To make more correct choices the fish had to follow the S+ and S- signals, rather than acquiring a side preference and reversing it. This experiment lasted for 25 days.

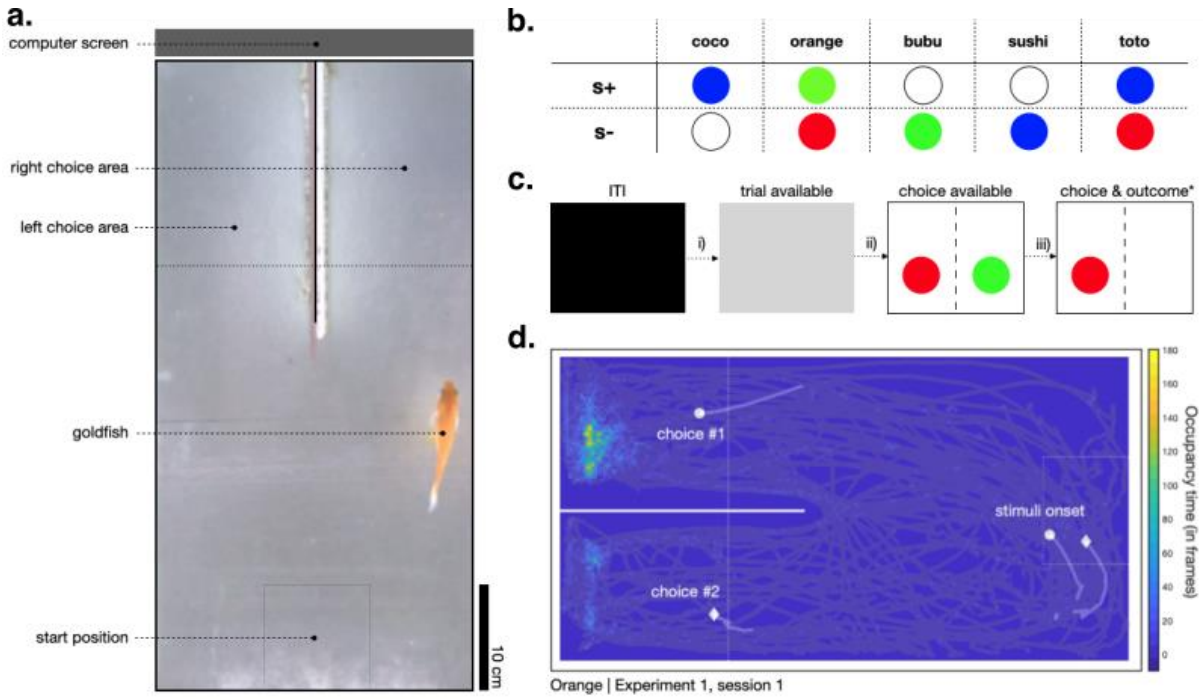


Figure 2. Goldfish were trained to associate colours with food rewards. a. Top view of the experimental tank, highlighting the start position and left and right choice areas. **b.** Stimuli colour allocation across subjects. **c.** Trial structure: Every trial started with an ITI, drawn from a uniform distribution (ITI duration, i): min = 20s, max = 40 s), which was signalled by a black screen. After this a trial became available the screen turned grey, signalling fish could move to the start position. The initiation time ii) was the time between a trial becoming available and a fish entering the start location. As soon as the fish entered the start location, 2 stimuli would appear on each side of the screen. Choosing S+ led to a pellet reward after a 5 second delay, choosing S- started a new ITI after a 5s delay. The response time iii) was the time between starting a trial and entering the left/right choice areas). **d.** Grey traces show the position tracks for an example animal and session. Markers depict the position of the fish (i.e., centroid) in the second leading to the stimuli onset and choice, for two example trials (shown with different markers), respectively. The underlying heatmap shows a 2d histogram of occupancy times (in number of frames) for the entire session. Same configuration as presented in a., rotated -90˚ for presentation purposes.

Data analysis

Real-time video tracking (see *Behavioural task control*) was used to control task contingencies and also generated a timestamped event list for each session. Preference and movement time (i.e., initiation and response times) data were derived from these event lists and analysed using custom Matlab code (R2020a, Mathworks) available at

https://github.com/PTMonteiro/GoFish_Ajuwon_etal_2022. Statistical analyses were conducted in RStudio (v1.2.5033; *The R Project for Statistical Computing*, 2018). For statistical analyses, choice proportion data was arcsine square-root transformed to normalise the residuals. One sample, one-sided t-tests against 50% were used to assess performance at group level.

In both experiments repeated measures ANOVAs were conducted to assess the effect of session (to detect learning effects). In *Experiment 2* repeated measures ANOVAs were also conducted to assess the effect of session terciles on trial initiation times and choices (to detect within session satiation or warming up effects). A type-1 error rate of 0.05 was adopted for all statistical comparisons.

Ethics statement

All experiments were conducted at the John Krebs Field Station and approved by the Department of Zoology Ethical Committee, University of Oxford (Ref. No. APA/1/5/ZOO/NASPA/Ajuwon/Goldfish), and were carried out in accordance with the current laws of the United Kingdom. Animals were cared for in accordance with the University of Oxford's "gold standard" animal care guidelines. All experimental methods were non-invasive. No food restriction was necessary as fish were fed high palatable pellets during daily-experimental sessions, supplemented by the end of the day in case fish did not eat the minimum daily requirements, and with raw spinach at the end of the last weekly experimental session. Their diet also included blood worms on weekends. Maintenance and experimental protocols adhered to the Guidelines for the Use of Animals in Research from the Association for the Study of Animal Behaviour/Animal Behavior Society ("Guidelines for the Treatment of Animals in Behavioural Research and Teaching," 2006). On completion, the fish were reintroduced into holding tanks and eventually returned to the supplier.

Results and Discussion

We illustrate the potential of *GoFish* for use in automated, closed-loop behavioural experiments with two discrimination learning experiments with goldfish.

In Experiment 1, fish (i) controlled the flow of trials by swimming to a start location, which triggered the onset of visual stimuli in two target sites, and (ii) expressed a choice by swimming to either target ROI, which triggered (or not, depending on choice) a food reward, followed by an intertrial interval, at the end of which the 'start' ROI became receptive and a new trial could be started. Multiple-trial sessions took place without intervention of the experimenter. This protocol was used in an acquisition and a reversal phase. The experiment and its results (smooth, significant acquisition and reversal, Figure 3a) are similar to those carried out by Kuroda et al. in zebrafish (Danio rerio; Kuroda et al., 2017). Repeated measures ANOVAs with session as the independent variable confirmed a significant increase in preference for the rewarded side in both the acquisition ($F_{4,16} = 3.02$, $P < 0.05$), and reversal ($F_{6,24} = 13.84$, $P < 0.0001$) phases. In the last session of each phase, subjects' preference for the rewarded side was significantly above 50% (acquisition phase: 88% ± 0.04 (mean ± s.e.m.); one-

sample t4 = 5.41 $P < 0.01$, reversal phase: 79% ± 0.04; one-sample t4 = 5.97 $P < 0.01$).

In Experiment 2, reward location was randomised on a trial-by-trial basis so that the coloured stimuli and the spatial cues were no longer redundant, instead only the former were reliable signals for reward. At group level, fish readily learned to track the location of reward (Figure 3b). A repeated measures ANOVA with session as the independent variable confirmed a significant increase in preference for the side displaying the S+ stimulus ($F_{24,96} = 2.01$, $P < 0.01$). Data from the terminal session show that the average proportion of rewarded choices was 69% ± 0.09. This result was significantly above 50% (one-sample t4 = 2.18, $P < 0.05$) even though one of the five fish failed to learn, as shown in Figure 3b. Since the same subjects were used in both experiments, carry-over effects from Experiment 1 (where reward site was constant across) may have influenced acquisition of the random alternation protocol in Experiment 2.

In addition to choice data, we used a real-time tracking pipeline for automated detection and recording of fish entry into the regions of interest (start zone, left choice zone, right choice zone). The tracking data gives direct access to relevant behavioural metrics, such as trial initiation time (i.e., the time animals took to be detected in the start zone following ITI offset; Figure 2c - ii) and choice response

times (i.e., the time from starting a trial to entering one of the choice zones; Figure 2c - iii).

As a metric for learning and motivational changes, we compared initiation times between the first and last sessions of Experiment 2 (Figure 3c), but found no significant differences (paired t4 = -0.86, P = 0.44).

In addition to choice proportion, we measured choice response times. This variable can be extremely informative: in previous studies and protocols it has been found that response times on both single-option and choice trials can be at least as informative regarding preferences and choice mechanisms as choice proportions (e.g., Monteiro et al., 2020). Overall, we found no significant differences in response time between trials in which fish chose correctly or incorrectly (Figure 3d, first session: paired t4 = 1.61, P = 0.18, last session: paired t4 = 0.29, P = 0.44).

Finally, we explored whether the proportion of correct choices varied within sessions by checking for trends across terciles of sessions. Such effects can occur if there are 'warming up' or satiation effects. Once again, we found no significant effects either early or late in training as revealed by repeated measures ANOVAs with session tercile as the independent variable (Figure 3e, first session: F2,9 = 2.29, P = 0.16, last session: F2,9 = 0.096, P = 0.91).

In summary, as a proof-of-concept demonstration for *GoFish,* a fully automated, closed-loop, and open-source experimental platform, we show that goldfish can reliably learn to (i) self-initiate trials, (ii) associate a fixed location with reward (iii) reverse their preference when the rewarded location changes, and (iv) associate colours with reward contingencies. We also present temporal data because, although no significant effects were found in this sample study, they illustrate what can be measured and suggest strategies for analysis.

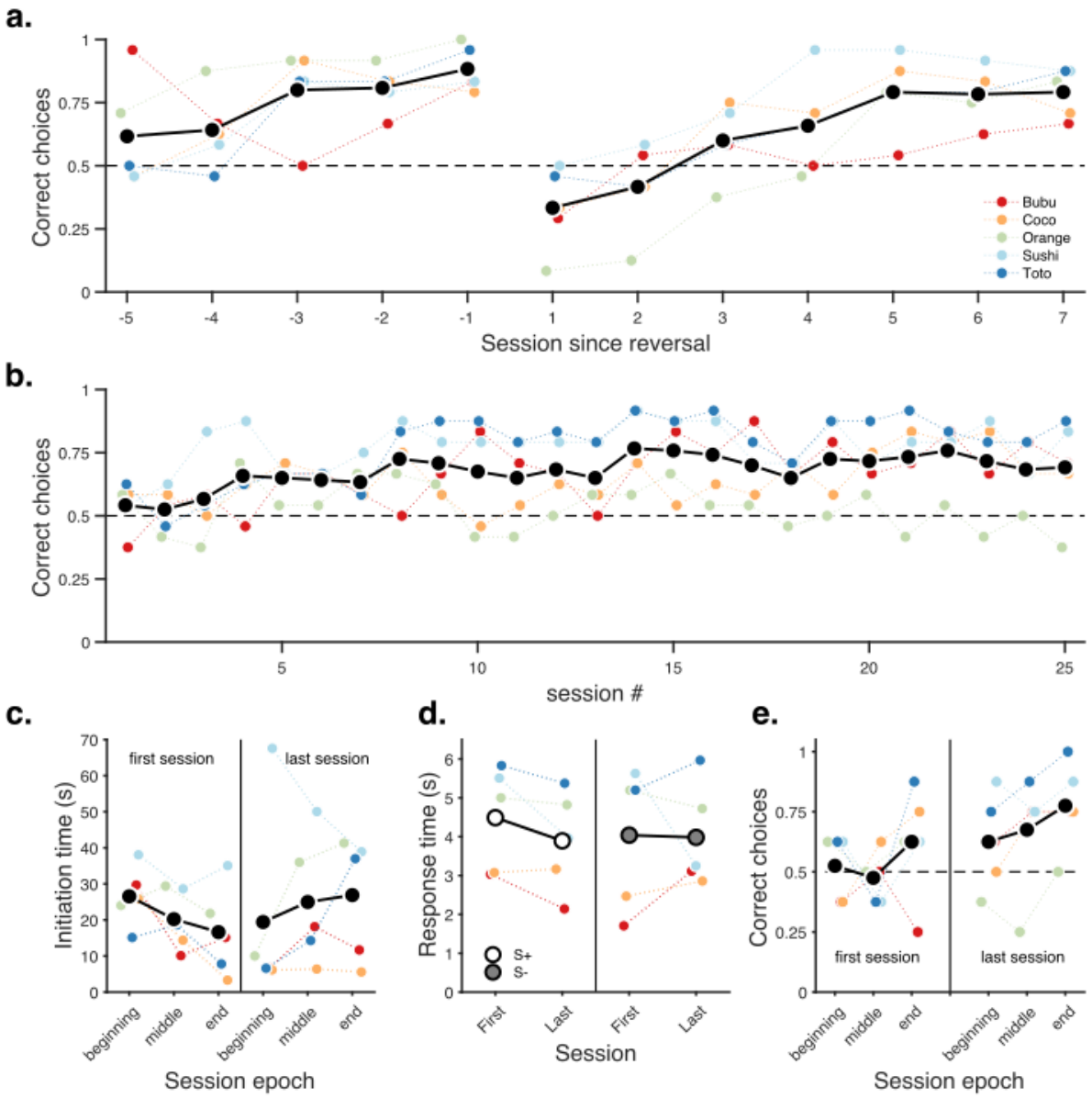


Figure 3. Goldfish learned a colour discrimination task with changing reward/cue/location requirements. a. Mean proportion of correct responses for *Experiment 1* during acquisition and reversal of a spatial discrimination. **b**. Mean proportion of correct responses during the visual discrimination task in *Experiment 2*. **c.** Initiation times for the first and last sessions of *Experiment 2*, split into session terciles. **d.** Response times towards S+ (left) and S- (right) stimuli for the first and last sessions of *Experiment 2*, split into the first and

last portions of the session, respectively. **e.** Proportion of choices as a function of session epoch (terciles), for the first (left) and last (right) sessions of *Experiment 2*, respectively. In all panels, black (white and grey, for panel d.) markers show group means and coloured markers the mean proportion of correct responses or median initiation or response times, respectively.

General Discussion

GoFish is a new platform for dynamic, fully automated behavioural experiments that facilitates high-throughput, highly reproducible research in fish or other aquatic organisms. *GoFish* is open-source, inexpensive, highly adaptable, and should be supported by a growing community of Bonsai users.

Critical to *GoFish*'s functionality is a novel reward pellet dispenser for which we provide design and assembly instructions, and Bonsai, the open-source programming language that is used to automate task contingencies and record data.

Using Bonsai in *GoFish* improves the user-friendliness of the system compared to proprietary experimental platforms for a number of reasons. Bonsai is free and compatible with a vast range of hardware devices meaning that users can easily source components cheaply or use already existing ones. Critically, Bonsai is a visual programming language, meaning that users with little or no previous coding experience can quickly develop effective workflows for task control and data analysis. In order to adapt our workflow for different protocols, users will need to learn the basics of Bonsai, which can be done through the extensive

documentation that exists on the Bonsai website, including example workflows and video tutorials.

As a generic experimental platform, *GoFish* provides advances and improvements over more common experimenter-controlled setups currently used in research on fish behaviour and cognition. These improvements have benefits in four domains: (i) methodology, (ii) animal welfare, (iii) reproducibility, and (iv) education.

Methodologically, the platform reduces the potential for unintended bias in experimenter-run tests, which are harder to run blindly. Also, having fully automated tasks reduces the chance of human errors. Moreover, in combination with the automation, the low cost of *GoFish* (Table 1) opens the possibility of testing multiple animals in parallel. Such standardisation across setups and subjects increases efficiency and helps to reduce inter-individual variability, ultimately contributing to a general refinement of procedures.

Methodological refinements will likely result in the reduction of the numbers of experimental animals used. Moreover, eliminating the experimenters´ presence during data collection reduces noises, shadows or other uncontrolled environmental changes thereby reducing subjects' stress levels, and improving subjects' welfare.

GoFish improves reproducibility, due to standardisation, and highlights the importance of low-cost, open-source tools for the advancement of scientific research.The fact that all components (including software) are open-source should afford further community-based system refinements over the long-term, enabling easier automated extraction of a wider range of behavioural metrics which should enrich the description of behaviour.

It is worth noting that studies have raised concerns regarding the applicability of automated operant training methods for fish, as studies with guppies have shown that automating procedures could lead to slower, unreliable, and task dependent outcomes compared to manually implemented tasks (Gatto et al., 2021). We hope our system, due to its flexibility, would enable us and others to explore this matter further.

The experimental configuration that we present here - a Y-maze setup for two alternative forced choice reversal learning and colour discrimination tasks is used as a proof-of-concept; *GoFish* is highly adaptable and could be used without configural changes in multiple other experimental paradigms e.g., quantity discrimination experiments (Potrich et al., 2022; Schluessel et al., 2022), behavioural timing (Talton et al., 1999), foraging (Aw et al., 2009; Newport et al., 2021), object recognition (Newport et al., 2016), and navigation (Burt de Perera & Holbrook, 2012). *GoFish* could also be used to implement experiments

using a range of set-ups differing to that reported here (e.g., open field and maze configurations that could employ a greater number of screens and/or feeders than we have). It is also worth noticing that *GoFish* could be used to present stimuli, in other sensory modalities: instead of using computer screens for visual stimuli presentation, Bonsai affords a large pool of interaction possibilities (e.g., adding a range of sensors and/or actuators, and sound libraries for auditory stimulus generation (Lopes et al., 2021)). Moreover, within Bonsai's framework our tracking routine based on colour thresholding could be extended to implement markerless (Kane et al., 2020 - https://github.com/bonsai-rx/deeplabcut) and multi-animal tracking (Guilbeault et al., 2021; Pereira et al., 2022 - https://github.com/bonsai-rx/sleap). Furthermore, our automatic pellet dispenser could easily be modified to use other regular-shaped rewards by laser cutting a different reward disk (Figure 1c; see also Arce & Stevens, 2022; Oh et al., 2017). With the present dimensions, the maximum number of rewards between re-fills is 40, which may be limiting for some applications. However, this number depends on the size of individual rewards, which may vary depending on the particular application of the feeder.

Finally, we note that the low price and scalability of the system makes it suitable for hands-on practical experiments and projects in education contexts (e.g., undergraduate projects, summer courses). It could be used for teaching basic

animal learning, experimental methods for behavioural research, and data processing (i.e., video tracking) and visualisation.

GoFish is a fully integrated, adaptable platform designed to facilitate the implementation of complex behavioural protocols in aquatic species. We hope that our platform accelerates the pace of refined behavioural research in a range of species that otherwise have been relatively underutilised in comparative and cognitive research programmes.

Author List: Victor Ajuwon, Bruno F. Cruz, Paulo Carriço, Champalimuad Foundation Scientific Hardware Platform, Alex Kacelnik, and Tiago Monteiro

Author contributions: VA, AK, and TM conceptualised and designed the experiments; VA and TM built the setup; VA, BC and TM wrote the *Bonsai* workflow; VA and TM collected the data; VA, BC, AK and TM analysed the data; PC and CRSHP designed and built the pellet dispensers; VA and TM wrote the first draft of the paper; all authors edited and reviewed the manuscript.

Chapter 4. Exploring non-instrumental information-seeking in goldfish

Abstract

Members of various species show strong preference for mixtures of positive and negative predictive signals (i.e., stimuli predicting delayed reward or absence of reward) over non-predictive alternatives, even if information does not improve outcomes, a phenomenon known as 'paradoxical choice'. This may be driven by an intrinsic reinforcing value of resolving uncertainty, or by secondary conditioning: if signals for sure reward (good news) are more reinforcing than signals for sure no-reward are inhibiting, equiprobable mixtures of both can have net reinforcing value. This preference has been shown so far in birds and mammals, but has not yet been investigated in other taxa, meaning we have an incomplete understanding of how the behaviour evolved, and the generality of the underlying mechanisms. We investigate whether paradoxical preference for informative signals is present in a fish, in the hope of expanding the taxonomic database while seeking potential dissociations between conditioning processes and information bias. Goldfish (*Carassius auratus*; n = 8) chose between two alternatives, both offering a 50% chance of reward after a 5s delay. Responding to the 'informative' alternative caused the immediate onset of a stimulus that anticipated the trial's outcome (reward/no reward), while choosing the 'non-informative' option triggered a stimulus that was equiprobably paired with either outcome. Subjects learned the contingencies of the stimuli, but in contrast to

mammals and birds, showed no preference between the informative and non-informative alternatives. Absence of evidence of a bias is not equivalent to evidence for absence, but these results raise the possibility that though conditioning mechanisms are ubiquitous across vertebrates, they may not always be sufficient to generate information biases, which appear restricted to certain taxa.

Introduction

In a world where uncertainty is pervasive, informative signals enable organisms to exploit opportunities, increasing the procurement of essential biological commodities such as food, or reducing the probability of undesirable risks. The fact that information is valuable inasmuch as it can lead to some behavioural change is part normative views of learning (for a review see Matute et al., 2019) and decision-making, in a range of fields including microeconomics (Mas-Colell et al., 1995), neuroeconomics (Rangel et al., 2008) and behavioural ecology (Stephens and Krebs, 1986; Vasconcelos et al., 2017). Within biology, these normative models share the assumption that seeking information is selected for because it causes behavioural adjustments that yield gains in substantial commodities such as food rewards. However, from a mechanistic perspective, reducing uncertainty may well be reinforcing *per se*, regardless of whether increased ability to predict future events correlates with substantial gains.

Over the last few decades, experimental observations of paradoxical decision-making with respect to reward maximisation have highlighted that individuals may ascribe an *intrinsic* subjective value to information, independent of any associated reward gains (Cervera et al., 2020; Gottlieb and Oudeyer, 2018; Kaplan and Oudeyer, 2007; Kidd and Hayden, 2015). In particular, in an

experimental paradigm often referred to as ‘paradoxical choice’, humans (Bennett et al., 2016; Brydevall et al., 2018; Iigaya et al., 2016; Kobayashi et al., 2019) and a range of mammal (Ajuwon et al. 2022; Bromberg-Martin and Hikosaka, 2011, 2009; Cunningham and Shahan, 2019; Ojeda et al., 2018) and bird (Fortes et al., 2018; González et al., 2020; McDevitt et al., 2022; Smith et al., 2016; Vasconcelos et al., 2015) species show a strong preference for advanced *non-instrumental* information about forthcoming rewards, even though in the lab the information cannot be used to improve beneficial outcomes, and may come at considerable cost. Subjects are typically presented with two options which deliver food or liquid rewards probabilistically, a fixed delay after the choosing response. In the informative option (Info) a stimulus is immediately presented after choice, eliminating outcome-related uncertainty, while in the uninformative option (NoInfo), outcomes are uncertain until the delay has lapsed. Critically, subjects cannot exploit the information provided in Info, which offers no more objective reward than NoInfo. Hence, some authors argue that preference for the informative option reflects intrinsically motivated information-seeking mechanisms concerning future events of biological significance.

Work in a range of vertebrate model organisms has been key to advancing our mechanistic, and evolutionary understanding of non-instrumental information-seeking behaviour, though fundamental questions remain. From a mechanistic perspective, neuroscientific studies on macaques (e.g., Blanchard et al., 2015;

White et al., 2019) have helped to advance the notion that non-instrumental information-seeking in animals is an analogue (or homologue) of human curiosity – an intrinsic desire to investigate and learn – reflecting psychological mechanisms aimed at resolving uncertainty (Bennett et al., 2016; Kobayashi et al., 2019; Kreps and Porteus, 1978) or filling 'information gaps' (Loewenstein, 1994). This view is supported by evidence in both humans (Brydevall et al., 2018; Charpentier et al., 2018) and monkeys (Bromberg-Martin and Hikosaka, 2011, 2009) showing that the same midbrain dopamine neurons can encode both information gain and primary rewards such as food.

Studies on pigeons have helped to characterise a range of experimental factors that influence non-instrumental information-seeking (Dinsmoor, 1983; González et al., 2020; McDevitt et al., 2016; Zentall, 2016), and have advanced an alternative mechanistic account of observed preferences that do not posit a sensitivity to information *per se*. According to this account, Pavlovian stimuli announcing positive outcomes become appetitive, conditioned reinforcers of behaviour, while aversive stimuli announcing negative outcomes are weakly inhibitive or ignored all together (Cunningham and Shahan, 2018; Dinsmoor, 1983; McDevitt et al., 2016; Smith et al., 2016). Similarly, a study on humans has suggested that 'good news' stimuli derive value by allowing subjects to hedonically savour forthcoming outcomes (Ligaya et al., 2016). From a functional perspective, some have suggested that such preferences are

maladaptive consequences of the psychological mechanisms responsible for 'sub-optimal' behaviour and may reflect unavoidable cognitive constraints that are also implicated in pathological human gambling (Laude et al., 2014; Ligaya et al., 2016; Zentall et al., 2019; Zentall and Stagner, 2011). However, this would imply that the mechanism producing the behaviour is maladaptive even in ecological contexts. This hypothesis has a cost, because it opens the puzzle for why such a robust (maladaptive) behavioural mechanism would not be eliminated and replaced by natural selection.

A study examining paradoxical choice in starlings has helped to shed light on the functional significance of preferences in the paradoxical choice task, by emphasising the distinction between the laboratory and the natural environment in which behaviour evolved. A foraging analysis by Vasconcelos et al. (2015; see also Fortes et al., 2017, 2016) indicates that preferences for advanced information leading to submaximal outcomes in the lab, would actually maximise reward rates in natural situations, because information pertaining to rewards will most often be useable by subjects, unlike in the artificial experimental protocol. Thus, paradoxical choice reflects adaptive mechanisms evolved to surmount common foraging challenges (Vasconcelos et al., 2018). The relatively recent discovery of robust non-instrumental information-seeking behaviour in rats (Ajuwon et al., 2022; Cunningham and Shahan, 2019; Ojeda et al., 2018) supports this notion: the prevalence of the phenomenon in distantly related taxa implies that non-

instrumental information-seeking is adaptation to common, widespread ecological challenges.

Though progress has been made in understanding the mechanistic and functional basis of non-instrumental information-seeking, so far studies have been restricted to a relatively small set of mammal and bird species, primarily in macaques and pigeons, but also rats and starlings. This means that we have an incomplete understanding of the evolutionary origin and function of this intriguing behaviour, and it is also not clear whether the mechanisms underlying non-instrumental information-seeking reflect universal properties of decision-making mehcnaisms in animal brains. Determining the ubiquity of such preferences in a broader range of species will help to develop more accurate and general theories of value-based decision-making that reflect known mechanisms of behaviour, and should also help to elucidate the neural substrates required for subjective information-value.

We address these issues here by examining non-instrumental information-seeking in a fish (goldfish; *Carassius auratus)* for the first time, using the paradoxical choice paradigm, and a novel experimental platform for closed-loop experiments on aquatic species (Ajuwon et al., 2023). The use of *GoFish* compared to experimenter-controlled set-ups common in fish studies, brings a similar level of automation and control found in other investigations of non-instrumental

information-seeking to our study, aiding interspecies comparisons of behaviour, while eliminating effects of experimenter bias.

Methods

Subjects

Eight goldfish ranging in size between 7 - 10cm, (age and sex unknown) participated in the current study. Animals were obtained from a local, commercial supplier (Goldfish Bowl, Oxford, UK).

Fish were housed in groups of two or three, in holding aquaria (60 x 35 x 31 cm; *length* x *width* x *height*) where they had access to a rock shelter, pebbles and artificial plants. They participated in experiments 5 times a week on weekdays and were fed *Fancy Goldfish Sinking Pellets* during experiments. This diet was supplemented with spinach following experiments on the last day of the week and bloodworms the day after. Fish were kept under a 12:12 h light: dark cycle using fluorescent lights. Water was maintained at a minimum of 21℃ using an internal heater and independent thermometer (pH: 8.2; ammonia: 0 ppm; nitrite: 0 ppm; nitrate: max. 30 ppm). Partial water changes were conducted at the end of each week and internal filters were cleaned every month. Each holding tank was aerated using an air pump.

For each daily session, fish were transported in a plastic jug to their experimental tank and then back to their holding tank at the end of the session. At the start of each day ~20L of water from all holding tanks were transferred to the experimental tanks in order to keep the environmental conditions as constant as possible. The experimental tanks were cleaned at the end of each week. All animals had different levels of previous experience in unrelated protocols, with 5 of the 8 individuals having additional experience in automated instrumental learning tasks.

Apparatus and task control

Experiments were run using *GoFish*, an open-source experimental platform for closed-loop behavioural experiments on aquatic species (Ajuwon et al. 2023). The full specifications of the *GoFish* platform, as well as task control information can be found in the source article, but for clarity we briefly outline the details here. The experimental set-up constituted a rectangular tank (60 x 30 x 36 cm; *length* x *width* x *height*) with 15 cm of water, a 17'' LCD computer screen (1920 x 1080; 60 Hz) for stimulus presentation, two custom-made pellet dispensers placed either side of an opaque acrylic divider in a y-maze configuration, an overhanging USB camera (1280x720 resolution), and an aquarium light (**Figure 1a**). Each reward delivery consisted of one *Fancy Goldfish Sinking Pellet.*

Experiments were run in parallel in two experimental tanks with 4 fish randomly assigned to each at the start of the experiment.

Task control was fully automated and implemented using a custom *Bonsai* workflow (Lopes and Monteiro, 2021). Task contingencies (i.e., stimulus presentation and reward delivery) were dependent on fish being within specified regions-of-interest (ROIs) in the experimental tank and were controlled using real-time video detection of fish movement. Within the tank three ROIs were defined: a 'start zone', 'left choice zone' and 'right choice zone'. These zones were not defined by physical boundary markings but were delineated on the video feed corresponding to fixed areas within the experimental tank. To detect fish presence in the ROIs a colour thresholding method was used. Video frames from these regions were recorded and converted to HSV colour space. A HSV range was applied to successfully detect the fish. The pixels of the resulting binarized frames from each ROI were summed continuously. Fish entries into the zones were recorded when the summed ROI pixel value exceeded a pre-set threshold (ROI thresholds were adjusted to each subject prior to the onset of the experiment). Thus, subjects were able to control the progress of trials and make choices via locomotion.

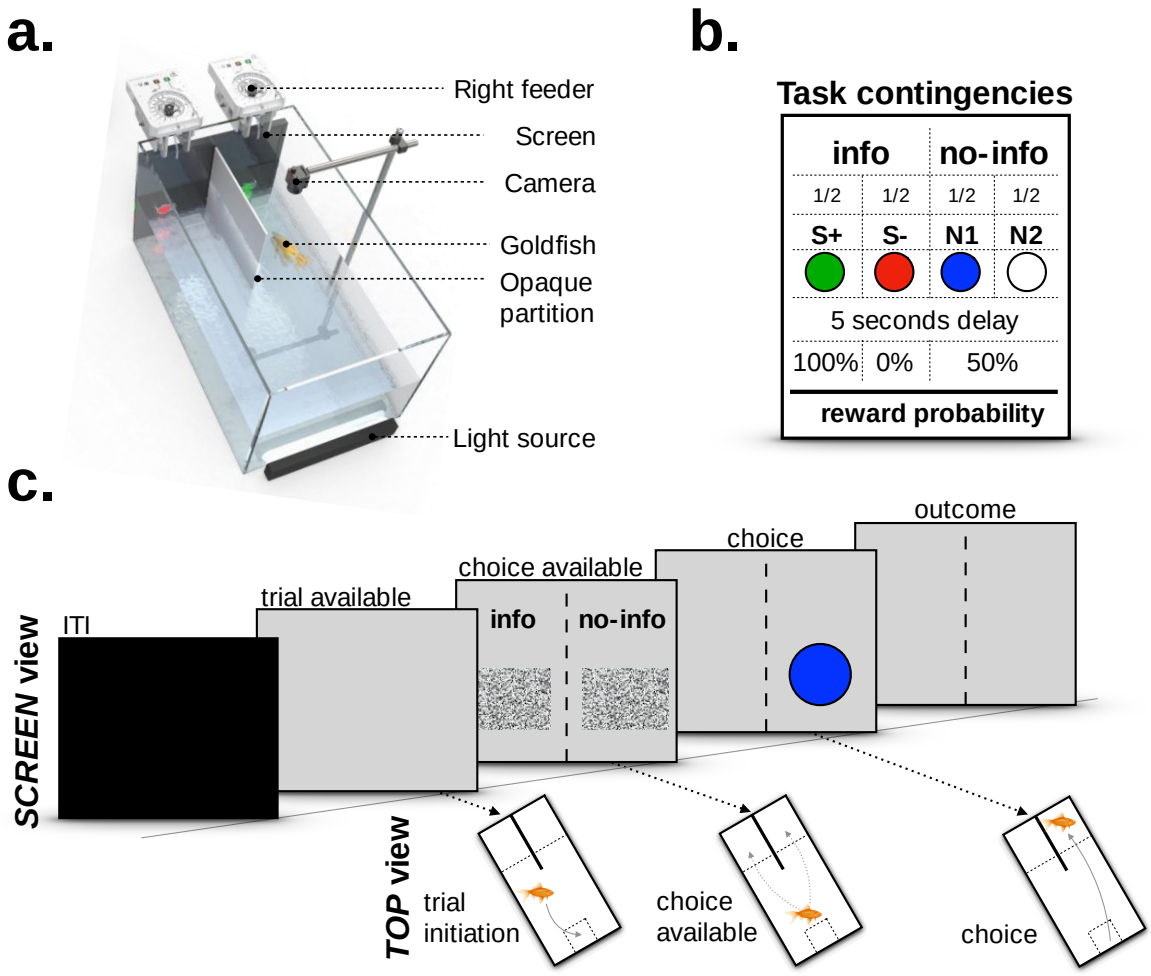


Figure 1. Exploring non-instrumental information-seeking in goldfish. a. The *GoFish* setup includes a rectangular tank (60 x 30 x 36 cm; length x width x height) with 15 cm of water, a 17'' LCD computer screen (1920 x 1080; 60 Hz) for stimulus presentation, two custom-made feeders placed either side of an opaque acrylic divider in a y-maze configuration, an overhanging USB camera (1280x720 resolution), and an aquarium light. **b**. Each subject chose between two options delivering reward probabilistically 5s after being chosen. There were two reward predictive cues in the informative option: S+ (100% reward probability) and S− (0% reward probability), and two cues in the non-informative option: N1 and N2 (both with 50% reward probability). **c**. Example trial structure. Trials were preceded by an inter-trial-interval (ITI; drawn from a uniform distribution: min = 20 s; max = 40 s) where the screen was

black, and behaviour had no consequences. The end of the ITI was signalled by a grey screen and from this moment on entering the start position would trigger the presentation a white noise stimulus in both options on choice trials, and in one of the options in forced trials. Subsequently swimming into one of two choice zones led to the presentation of a visual stimulus followed by the corresponding outcome after a 5s delay, and then the onset of the ITI. Task contingencies (i.e., stimulus presentation and reward delivery) were dependent on fish being within 3 specified regions-of-interest (ROIs) in the experimental tank (i.e., 'start', 'left choice' and 'right choice' zones). Task control and real-time video analysis of fish movement were fully automated and implemented using a custom *Bonsai* workflow.

Stimuli

Within the main experiment and training there were four reward-predictive cues: coloured circles presented on a 17'' LCD computer screen (1920 x 1080; 60 Hz) all with a duration of 5s and each associated with a reward probability (**Figure 1b**). For each subject there were two cues for the informative option: S+ (100% reward probability) and S− (0% reward probability), and two cues for the non-informative option: S1 and S2 (both with 50% reward probability). These were red, green, blue, and white circles, 3.5 cm in diameter on a grey background, presented with centers positioned 5 cm from the bottom of the tank and 7 cm from each side wall. We chose colours that have been physiologically (Neumeyer, 1984) and behaviourally (Ajuwon et al., 2023; Zerbolio and Royalty, 1983)

proven to be discernible by our experimental species. Assignment of colours to reward probabilities was partially counterbalanced across subjects to avoid the possibility of option preferences being influenced by any intrinsic aversive or attractive properties of the stimuli.

In the main experiment we used a monochromatic white noise rectangle (13.5 x 12 cm, Gaussian: mean=0, variance=10) to indicate whether an arm was available for a choice after a trial was initiated. In the pre-training phase, the white noise stimulus was used to signal that reward delivery was contingent on fish swimming to a specific location in the tank.

General Procedure

Real-time video detection of subjects was used to control task contingencies, including the delivery of food rewards and the presentation of visual stimuli on a screen. In experiment 1 goldfish were repeatedly presented with a choice between two alternatives of equal average profitability (expected reward/ expected time) delivering food rewards probabilistically after a 5s post-choice delay. In the informative option (Info) visual stimuli (S+ / S-) anticipated the delayed outcomes immediately after choice, while in the non-informative option (NoInfo), outcomes were uncertain (N1/ N2) until the delay period lapsed. In

experiment 2, goldfish were presented with a series of choices between reward-predictive stimuli that they had encountered in experiment 1, to determine their knowledge of task contingencies.

Pre-training

Pre-training consisted of three phases: (i) experimental tank acclimatisation, (ii) choice zone training, for subjects to learn that swimming into either the left or right choice zone (outside of the ITI) could be reinforced, and (iii) start position training, for subjects to learn that a start position had to be entered before subsequent behaviour could be reinforced. Advancing through the phases depended on the individual subject's performance. For further details see 'Pre-training' in (Ajuwon et al. 2023).

Experiment 1: Paradoxical choice

We used a trial-based chain procedure as displayed in (**Figure 1c**). Every fish was presented with 2 daily sessions of 24 trials for 45 days. There were 2 kinds of trials: 2-option choice trials and 1-option forced trials. The proportion of each of these trials within a session changed as the experiment progressed (**Table 1**)

but where a session was composed of both types of trial, they were randomly intermixed.

Table 1. Proportion of trials across experimental sessions

Day	Session	Forced	Choices
1-15	1	24	--
	2	"	"
16-20	1	"	"
	2	16	8
21-40	1	"	"
	2	"	"
41-45	1	8	16
	2	"	"

Trials were preceded by an inter-trial-interval (ITI; drawn from a uniform distribution: min = 20 s; max = 40 s) where the screen was black, and behaviour had no consequences. The end of the ITI was signalled by a grey screen and from this moment on, entering the start position would trigger the presentation of the white noise stimulus in one of the two choice zones (1-option forced trial) or both choice zones (2-option choice trial). Swimming into a choice zone with the white noise stimulus present led to the disappearance of the white noise stimulus, the presentation of a visual stimulus for 5s and then a 50% chance of reward delivery.

There were two potential options for subjects to choose, which could either be presented alone (1-option forced trial) or simultaneously (2-options choice trial). Swimming into the choice zone of the informative option (Info) resulted with equal probability in either S+ for 5s, which was paired with reward, or S- for 5s, which was paired with no reward, thereby reliably informing the subject of the forthcoming outcome. Swimming into the non-informative (NoInfo) choice zone, on the other hand, resulted with equal probability in either of two cues: N1 or N2, which were both associated with a 50% probability of either outcome; therefore, neither cue informed the subject of forthcoming reward. The side of the Info/NoInfo option was consistent for each individual, but counterbalanced across subjects.

Experiment 2: Stimulus choice

The paradoxical choice experiment was followed by a second experiment that lasted for 5 days. In this experiment, rather than choosing between Info and NoInfo sides, subjects were presented with the following pairs of reward predictive stimuli to choose between: S+ vs N1 or N2, and S- vs N1 or N2. The aim of this experiment was to verify the extent to which subjects had learned to discriminate between the reward predictive stimuli in Info and NoInfo, which is critical for the interpretation of preferences in the paradoxical choice experiment.

There were 2 daily sessions of 24 trials per day. The order of stimulus choice presentations was randomised.

Trials were preceded by an inter-trial-interval (ITI; drawn from a uniform distribution: min = 20 s; max = 40 s) where the screen was black, and behaviour had no consequences. The end of the ITI was signalled by a grey screen and from this moment on, entering the start position would trigger the presentation of S+ or S- in one choice zone and N1 or N2 in the other. Swimming into a choice zone led to the reward outcome of the corresponding stimulus after a 5s delay and the onset of the ITI.

Fish Tracking

To track the fish centroid in real-time a colour thresholding method as described in Ajuwon et al. 2023) was implemented using a custom *Bonsai* workflow. Video was recorded at approximately 33 fps.

Data analysis

Data processing and analysis was carried out in MATLAB 2022b. Statistical analyses were conducted in RStudio (v1.2.5033; *The R Project for Statistical Computing*, 2018). To normalise the residuals in statistical analyses, entropy data from fish tracking in experiment 1 was square-root transformed and choice proportion data from experiment 2 was arcsine square-root transformed. A type-1 error rate of 0.05 was adopted for all statistical comparisons. A Bonferroni correction was applied for paired pairwise comparisons.

To examine whether subjects were able to discriminate the contingencies of each reward-predictive stimulus during experiment 1, we quantified fish spatial occupancy in a 4s-window preceding reward outcomes (i.e., 0.5s after choice and 0.5 pre outcome) using entropy (a measure of randomness) as a metric to parameterise the spatial distribution of subjects from an input image capturing this 4s window. Entropy is given by

where P is a matrix containing the normalized histogram counts returned from the MatLab imhist function on the input image. Since N1 and N2 have identical

reward contingencies the mean entropy score of both stimuli was calculated for each individual and used in subsequent analysis.

In order to further examine subjects' ability to discriminate the reward predictive stimuli, we quantified the strength of preference for S+ relative to S- in Experiment 2, by calculating the following metric for each subject during each day of experiment 2: Discrimination Metric$_{(S+ \text{ vs } S-)}$ = proportion of S+ choices/ (proportion of S+ choices + Proportion of S- choices).

Ethics statement

All experiments were conducted at the John Krebs Field Station and approved by the Department of Biology Ethical Committee, University of Oxford (Ref. No. APA/1/5/ZOO/NASPA/Ajuwon/Goldfish), and were carried out in accordance with the current laws of the United Kingdom. Animals were cared for in accordance with the University of Oxford's "gold standard" animal care guidelines. All experimental methods were non-invasive and no food restriction was necessary. Maintenance and experimental protocols adhered to the Guidelines for the Use of Animals in Research from the Association for the Study of Animal Behaviour/Animal Behavior Society ("Guidelines for the treatment

of animals in behavioural research and teaching," 2018). On completion, the fish were reintroduced into holding tanks and retained for further experiments.

Results

Goldfish spatial location shows discrimination of stimulus contingencies

The focus of this study was to investigate preference between two alternatives, Info and NoInfo, that differ only in their ability to signal forthcoming outcomes, therefore it was important to establish that subjects could discriminate between the reward predictive stimuli in each option. To this end, we examined goldfish locomotion during the delay period preceding reward outcomes after choice **(**0.5 s after choice and 0.5 s pre-outcome**)**, while reward-predictive stimuli were presented in Experiment 1. **Figure 2** shows strong evidence that goldfish could discriminate between cues: subject proximity to the sites of reward delivery during this period varied appropriately according to each cue's reward contingency. When reward was forthcoming (**Figure 2a,b** left) or potentially forthcoming (**Figure 2a,b** center) subjects generally remained in the choice zone, close to the pellet dispenser. When reward was not forthcoming (**Figure 2a,b** right) subjects actively swam away from the choice zone, and pellet dispenser, suggesting that S- may have acquired aversive properties.

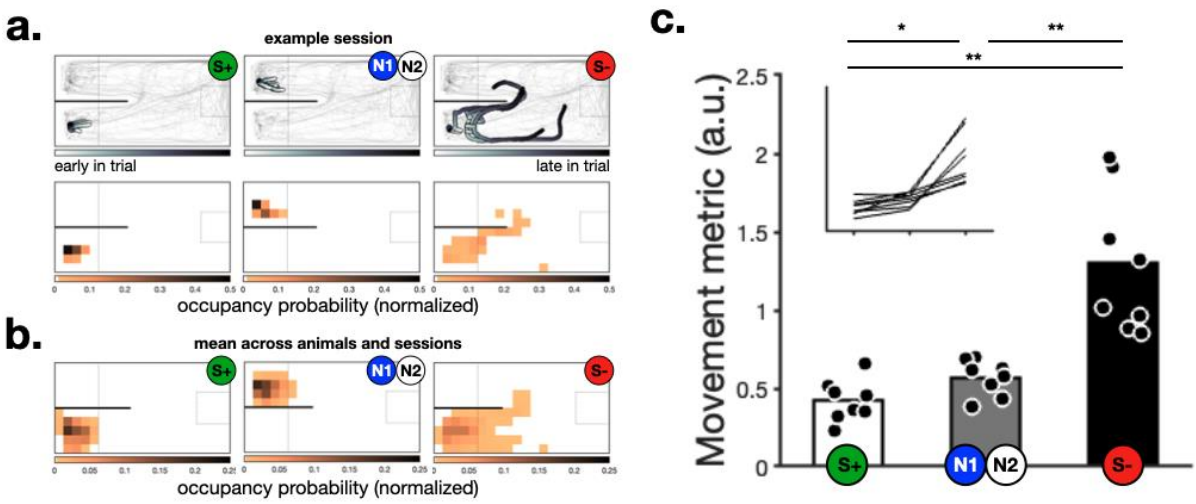


Figure 2. Goldfish spatial location varied according to cues' reward contingencies. a. Top row: Trajectories of a representative animal and session from Experiment 1 during a 4s post-choice window when reward predictive cues were presented (0.5s post-choice – 0.5s pre-outcome), split by trials in which the animal was presented with S+ (left), N1 or N2 (middle), or S- (right). Thin grey lines show the trajectories for the entire session for reference. Middle row: Corresponding normalised 2D histograms of fish occupancy. **b.** Same as above but averaged across all animals and sessions (n=8 animals, 5 sessions each). **c**. Movement metric computed as the entropy of subjects spatial distributions, where larger values correspond to more movement. Markers show individual animal means and bars the group averages. Inset shows individual trends.

We quantified behavioural responses by calculating the entropy of subjects' spatial probability distributions during a 4s-window preceding reward outcomes (i.e., 0.5s after choice and 0.5s pre outcome). Higher entropy values indicate subjects occupied a larger portion of the experimental tank during this period. Data pooled from the last 5 days in **Figure 2c** shows that fish spatial occupancy during cue presentation corresponded to the reward probability associated with the cue. The entropy of spatial distributions was lowest when S+ was presented (0.42 ± 0.047; mean ± s.e.m.), higher when N1/N2 were presented (0.57 ± 0.041) and greatest when S- was presented (1.3 ± 0.16), reflecting that subjects spent more time nearer the site of reward delivery when reward probability was greater. A one-way repeated measures ANOVA on these data with stimulus as a within-subject factor and entropy as the response variable revealed a significant effect of stimulus (S+, N1/N2, S- ; $F_{2,14} = 27.48$, $P < 0.0001$). Post-hoc pairwise comparisons showed that entropy was significantly different between all stimulus pairs (S+ vs N1/N2, $P < 0.05$; S- vs N1/N2, $P < 0.01$; S+ vs S-, $P < 0.01$), indicating that subjects accurately ranked stimuli in order of reward probability.

Goldfish choice is modulated by stimulus outcomes

To further confirm that subjects were tuned to the reward contingencies of each cue, in Experiment 2, over 5 days, we presented subjects with pairswise choices between stimuli from each alternative: S+ vs N1/N2 and S- vs N1/N2. If, as assumed, subjects had learned the outcomes associated with each cue in Info, we expected that they would show a greater preference S+ over N1/N2 and N1/N2 over S-.

Consistent with results from Experiment 1, preferences pooled over the 5 days of Experiment 2 showed that subjects discriminated between S- and N1/N2 (S- vs N1/N2; 17%, ± 0.029; t_7 = -8.97, P < 0.0001). However, the results from Experiment 2 do not explicitly show evidence of discrimination between S+ and N1/N2 in choices. Even though selecting S+ resulted in certain reward while N1/N2 only lead to a 50% chance of reward, goldfish showed no preference for S+ over N1/N2 (S+ vs S1/S2; 46%, ± 0.072 s.e.m.); choices between S+ and N1/N2 did not deviate significantly from that expected by chance (50%; t_7 = -0.63, P = 0.551). This is inconsistent with a reward-maximizing preference, showing that valuation of a 50% chance of rewards was sufficient to sustain an approach response as strong as that of S+ signalling certain reward. Given that subjects displayed the ability to discriminate between all cue pairs in Experiment

1, the explanation for this indifference must respond to reasons other than lack of discrimination.

Figure 3 (left) shows that subjects made a higher proportion of choices for S+ than for S- when presented against either of the NoInfo stimuli (See Figure S1 for individual subject data). A two-way repeated measures ANOVA on data across all 5 days of Experiment 2, with stimulus and day as within-subject factors and proportion of choices for S+/S- as the response variable revealed a significant effect of stimulus (S+ or S- ; $F_{1,35} = 85.19$, $P < 0.0001$), but no significant effect of day ($F_{4,28} = 0.87$, $P = 0.0494$), and no significant interaction ($F_{4,35} = 1.13$, $P = 0.357$). To determine the relative preference for S+ over S-, for each individual we calculated a Discrimination Metric$_{(S+ \text{ vs } S-)}$ (see methods), where values >0.5 indicate a greater preference for S+ and values <0.5 indicate a greater preference for S-. **Figure 3** (right) shows that pooled over the 5 days, each individual had a discrimination metric above 0.5. The mean Discrimination Metric$_{(S+ \text{ vs } S-)}$ across all individuals was significantly greater than 0.5 (0.73, ± 0.036; mean ± s.e.m.; $t_7 = 6.53$, $P < 0.001$). Taken together, the results from Experiment 2 indicate that subjects could correctly discriminate between, and accurately rank S+ and S-.

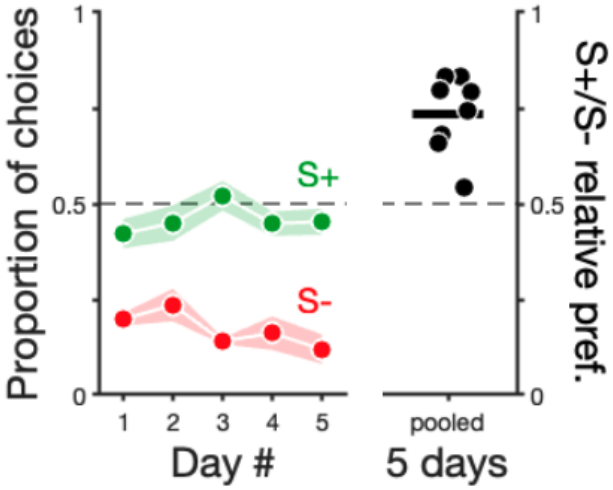


Figure 3. Goldfish preferences between reward-predictive stimuli in paradoxical choice. Rather than choosing between Info and NoInfo, in Expriment 2 subjects were presented with the following pairs of reward predictive stimuli to choose from: S+ (100% reward) vs N1 or N2 (50% reward), and S- (0% reward) vs N1 or N2 (50% reward). Left: Proportion of choice for S+ (green) or S- (red). Markers show group means and shading the s.e.m. Right: Relative preference for S+ over S- (proportion of S+ choices/ (proportion of S+ choices + proportion of S- choices) pooled over Experiment 2. Markers show individual means, and line shows group average (n= 8 animals).

Goldfish show no preference for advanced information in paradoxical choice

Our key result is that though subjects showed adequate knowledge of the contingencies in the paradoxical choice task, in contrast to a range of mammal and bird species, goldfish show no evidence of preference for advanced non-instrumental information (**Figure 4**). Over 30 days subjects were presented with a choice between Info, an informative option in which delayed, probabilistic reward outcomes were signaled immediately-post choice, and NoInfo, a non-informative option in which outcomes were uncertain until the post-choice delay period had lapsed. A one-way repeated measures ANOVA with day as a within-subject factor and proportion of choices for Info as the response variable showed no significant effect of day ($F_{29,203} = 0.99$, $P = 0.493$), indicating that the subjects did not develop a preference for either option. This was also reflected in data pooled from the last 5 sessions of the experiment, showing that preference did not deviate significantly from that expected by chance (45%, ± 0.057 s.e.m.; $t_7 = -0.74$, $P = 0.482$). Notice that in this protocol, because reward probability is the same in informative and non-informative options, the different species did differ in their experienced reward probabilities.

Consistent with previous studies (e.g., Monteiro et al., 2020) **Figure S2** shows a negative correlation between individuals' strength of preference for the informative option and the difference in response time between both options. In other words, across individuals we found a trend where greater preference for NoInfo is associated with shorter reaction times towards this option (Estimated model: $y = -0.03x + 0.51$; slope 95% CI = [-0.07 to 0.33]).

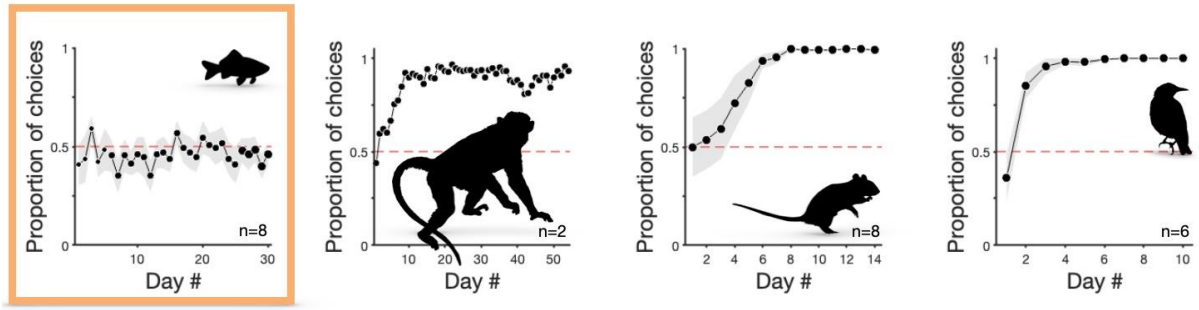


Figure 4. **Daily preference for the informative alternative across 4 different species in similar paradoxical choice protocols.** From left to right: Goldfish (*Carassius auratus*; n = 8 animals, 30 sessions each) from the present study. Rhesus macaque (*Macaca mulatta;* n = 2 animals, first 54 sessions). Adapted with permission from (Bromberg-Martin and Hikosaka, 2009). Rats (*Rattus norvegicus*, n = 8 animals, 14 sessions each). Adapted with permission from (Ajuwon et al., 2022) - *Figure 3*. European starlings (*Sturnus vulgaris*, n = 6 animals, 10 sessions each). Adapted with permission from (Vasconcelos et al. 2015) - *Figure 2*. Markers show group means and shading the s.e.m. In the left-most panel larger data markers reflect a greater number of choices per day and the orange box highlights data are from the present study. Animal silhouettes from https://www.phylopic.org/ under Creative Commons licenses.

Discussion

Our results show that in contrast to the vertebrate species previously tested (humans, monkeys, rats, pigeons and starlings), goldfish do not display a preference for advanced non-instrumental information pertaining to probabilistic reward outcomes in the paradoxical choice task (**Figure 4**), though they learnt task contingencies. Our results have implications for our understanding of the putative neural and psychological mechanisms underlying information-seeking behaviour, as well as its functional significance and evolutionary origins.

Key to the interpretation of our results is the ability of subjects to discriminate between the reward-predictive cues that critically distinguish the informative and uninformative alternatives. Two independent pieces of evidence indicate that the goldfish in our study could distinguish between the reward predictive stimuli, excluding the possibility that indifference at choice time was due to a putative inability to learn task contingencies. First, during the delay period between choice and outcomes when cues were presented in Experiment 1, subjects' proximity to the site of reward delivery showed that they treated the cues according to their reward contingencies: the greater the likelihood of signalled reward, the more time spent close to the site of reward delivery (**Figure 2**). Similar discrimination has been shown in rats that did show a preference for advanced information

(Ajuwon et al., 2022; Ojeda et al., 2018). Secondly, when given a direct choice between reward predictive stimuli in Experiment 2, subjects showed a preference for cues announcing 50% reward probability over those announcing no reward (**Figure 3**), which is similar to choices in starlings that showed a preference for the informative option (Vasconcelos et al., 2015). Curiously, in this experiment, subjects showed indifference between signals of sure reward and those for 50% reward. This highlights that in choices, the valuation of a 50% chance of reward was sufficient to sustain an approach response as strong as that of S+ signalling certain reward. This is inconsistent with rate-maximisation, but may reflect increasing satiation throughout a session. However, this does not preclude the overall conclusion that subjects could discriminate between cues, because behaviour in the first experiment indicated accurate discrimination.

The protocol that we employed to investigate non-instrumental information-seeking in goldfish was designed to retain as much as possible the critical features of the task as deployed in the other species, where a preference for advanced information had been observed. Colours that have been physiologically (Neumeyer, 1984) and behaviourally (Ajuwon et al., 2023; Zerbolio and Royalty, 1983) proven to be salient cues for goldfish were selected as reward predictive signals, and colour-reward contingency was counterbalanced across subjects, eliminating the possibility that potential preferences were masked by biases towards particular colours. Furthermore, unlike in a number of similar studies

where subjects incur reward losses by selecting the informative option (Fortes et al., 2016; McDevitt et al., 2018; Pisklak et al., 2019; Stagner and Zentall, 2010; Vasconcelos et al., 2015; Zentall et al., 2015), in our study both the informative and uninformative option were equally profitable. Therefore, unlike previous studies using rats (Martínez et al., 2017; Ojeda et al., 2018; Trujano et al., 2016; Trujano and Orduña, 2015), the lack of preference we observed cannot be explained by a sensitivity to reward losses incurred by selecting the informative option.

From a procedural perspective, task parameters other than reward profitability have been shown to influence the strength of preference for advanced information in paradoxical choice (for reviews see McDevitt et al., 2016; Zentall, 2016), these include the salience and modality of reward predictive stimuli (Ajuwon et al., 2022; Chow et al., 2017; Vasconcelos et al., 2015; Zentall et al., 2019), the length of delay between choice and reward outcomes (longer delays lead to greater *Info* preference: Cunningham and Shahan, 2019; Iigaya et al., 2016; McDevitt et al., 2018; Spetch et al., 1990), the proportion of forced trials subjects are presented with (more forced trials increase preference: McDevitt et al., 2022) and the action subjects must make to express preference (Dunn and Spetch, 1990; Kendall, 1974). Therefore, it is possible that the lack of preference for advanced information that we found in goldfish is only relevant to the specific combination of parameters that we implemented here, with a particular issue being that the

length of delay between choice and reward outcomes was too short to elicit a preference for the informative option. However, *a priori* it is not obvious why this would be the case, or indeed what other combination of parameters would elicit preference for the informative option, given that similar parameter sets have been sufficient to generate robust preferences in a range of other species. We designed our experiment so as to minimise the likelihood of this possibility. We used and counterbalanced stimuli to which goldfish are known to respond and selected a delay period (5s) long enough to elicit preferences in monkeys (i.e., >2.5s; Blanchard et al., 2015; Bromberg-Martin and Hikosaka, 2011, 2009) but shorter than the 10s used in rat and pigeon studies, to reduce the possibility of task disengagement in the freely swimming fish. Furthermore, our subjects underwent extensive pre-training in the form of forced trials, and forced trials were interspersed throughout the main experiment, eliminating the possibility that side biases prevented subjects from learning the contingencies of Info in the first place. These considerations, taken together with the fact that subjects learnt to discriminate the cues associated with each option under our set of experimental parameters, lead us to believe that our results reflect a genuine species difference between goldfish and other tested vertebrate models in sensitivity to, and preference for, advanced information about forthcoming outcomes.

This begs the question of why such species differences exist, which may be due to a number of reasons. From a mechanistic perspective, this could be because

goldfish lack sensitivity to reductions in uncertainty, or the ability to computationally quantify reward related uncertainty - which may be a product of the simpler neural architecture in goldfish than the mammal and bird species so far tested. This would be consistent with the finding that while advanced information about upcoming rewards is encoded in highly conserved brain regions such as the midbrain (Bromberg-Martin and Hikosaka, 2009) and lateral habenula (Bromberg-Martin and Hikosaka, 2011), more derived regions within the mammalian neocortex also represent gains in advacned non-instrumental information (Bromberg-Martin and Hikosaka, 2015; White et al., 2019). Therefore, goldfish might show lack of preference because they lack an analogue of the derived neural structures implicated in computations of information gain (and other cognitive processes), that are present in mammals (the neocortex) and birds (the nidopallium caudolaterale). If confirmed, this suggests that increased cognitive sophistication paradoxically leads to loss of foraging performance in some cases: species sophisticated enough to find information rewarding *per se*, perhaps because they can use it at a later time, may end up choosing subopimtally when information seeking is counterposed experimentally against reward probability.To test this proposal, information-seeking behaviour should be examined in other fish species: a lack of preference in a range of fish species with diverse ecologies would support explanations based on neural constraints.

Another explanation for lack of preference for advanced information in goldfish concerns conditioned reinforcement, which is an alternative mechanistic explanation for learned preferences for advanced information that does not rely specifically on sensitivity to reductions in uncertainty *per se.* This account argues that subjects select the informative option to generate 'good news' signals anticipating certain forthcoming rewards, which acquire reinforcing properties by association with outcomes of positive valence (Ajuwon et al., 2022; Dinsmoor, 1983; Roper and Zentall, 1999; Smith et al., 2016). In species that show preference for the informative option, this account posits that any putative inhibitory effects of signals for 'bad news' in the informative option are relatively small or ignored by subjects, so that overall, subjects prefer the informative option. Where rat preferences for the informative option have been shown to be qualitatively less robust than other species (Cunningham and Shahan, 2019; Martínez et al., 2017; Ojeda et al., 2018), it has been suggested that this is because they are more sensitive to conditioned inhibition through signals for bad news in the informative option (Alba et al., 2018; Martínez et al., 2017; Trujano et al., 2016). It is therefore possible that goldfish do not prefer the informative option because signals for bad news acquire inhibitory properties powerful enough to eliminate any reinforcing effects the signal for good news has. Some of our results indicate that this may be the case. In the main experiment, when presented with S-, subjects generally rapidly swam away from the stimulus (**Figure 2**), and in the second experiment subjects rarely selected S- when it was presented alongside

other stimuli (**Figure 3**) suggesting it has aversive properties which could inhibit Info preference. Manipulations involving the removal of S- will help to elucidate the extent to which conditioned inhibition plays a role in goldfish preferences for advanced information.

Functionally, the lack of preference for advanced information in goldfish may reflect a difference in foraging mode compared to other species so far tested. Vasconcelos et al. (2015) argued that preferences for a lower pay-off alternatives providing advanced non-instrumental information in the lab may reflect adaptive mechanisms that are rate maximising in natural circumstances, where information about forthcoming rewards will likely be useable.They envisaged that decision-makers could use such information to avoid unnecessary opportunity costs incurred while pursuing prey. However, for organisms like carp that are primarily suction feeders and scavengers, such information may not be as ecologically relevant. Therefore, in these cases sensitivity to information about forthcoming rewards may not have conferred a rate-maximising selective advantage, and may come at considerable cognitive cost.

Our results may also reflect differences in the ontogeny of individuals that have been tested in the paradoxical choice paradigm. Unlike starlings which were wild-caught (Vasconcelos et al., 2015), the goldfish in the present study were procured from commercial suppliers. Given the easy access to food and ideal living

conditions our goldfish presumably developed in, it is possible that during development they did not learn to seek non-instrumental information, as there were few marginal benefits associated with the behaviour, i.e., where food is made available in sufficient quantities there is little incentive to acquire and maintain curiosity-like behaviour within a lifetime. This is different for the starlings on the other hand, who in natural foraging scenarios may have benefitted from advanced information about food sources prior to lab experimentation, and therefore may have acquired a positive association with advanced information, leading to robust Info preference in experiments. This explanation however, can only partially account for our results because other domesticated species such as rats, pigeons and macaques do show a preference for the informative option, highlighting that other factors are likely at play.

Lastly, it is worth considering the evolutionary implications of our results. Our data alone do not allow us to definitely elucidate the origins of non-instrumental information-seeking behaviour, but together with further studies a clearer picture should emerge. It is tempting to speculate that based on the present results and parsimonious arguments, non-instrumental information-seeking was acquired by an ancestor common to the tetrapod lineage, which would explain its presence in humans, macaques, pigeons and starlings, but not goldfish. However, there are a number of other possibilities. Firstly, goldfish preferences might not be reflective of fish in general, in which case it could be argued that curiosity-like information

seeking is an innovation unique to the vertebrate clade as a whole, rather than the tetrapods in particular. One scenario for this would be that goldfish lost the ability to seek non-instrumental information, which is plausible given that unlike most fish species they have undergone hundreds of years of intensive artificial selection (Chen et al., 2020). Experiments on a broader range of fish species will help to disambiguate these proposals. Secondly, another possibility is that rather than being an ancestrally derived trait, non-instrumental information-seeking has evolved independently multiple times in the lineages in which it has been observed. One way to explore this is to investigate whether the same mechanisms are responsible for non-instrumental information-seeking in diverse taxa: this would strengthen the case for a shared origin. The present data cannot disambiguate these particular hypotheses, but open up a number of interesting avenues for research concerning the evolutionary origins of non-instrumental information-seeking behaviour.

Supplementary materials

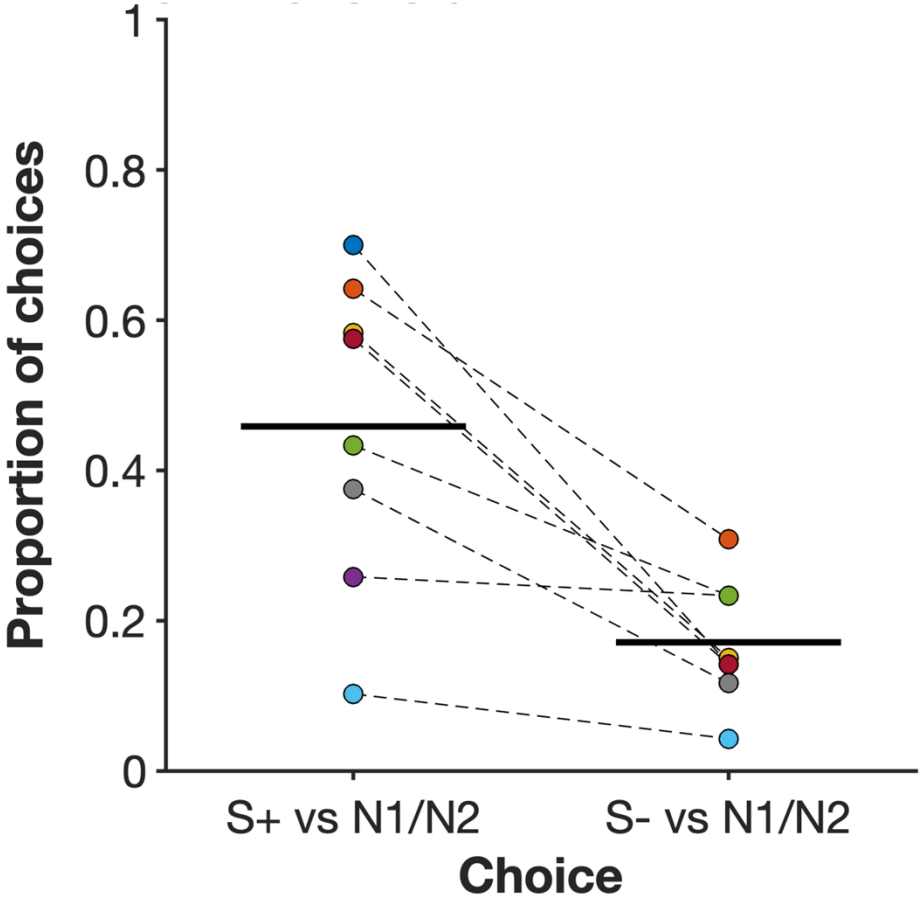


Figure S1. **Individual goldfish preferences between reward-predictive stimuli.** Mean proportion of choices pooled from the 5 days of experiment 2 for each individual. Horizontal bars represent group means and dashed lines connect data points belonging to the same subject.

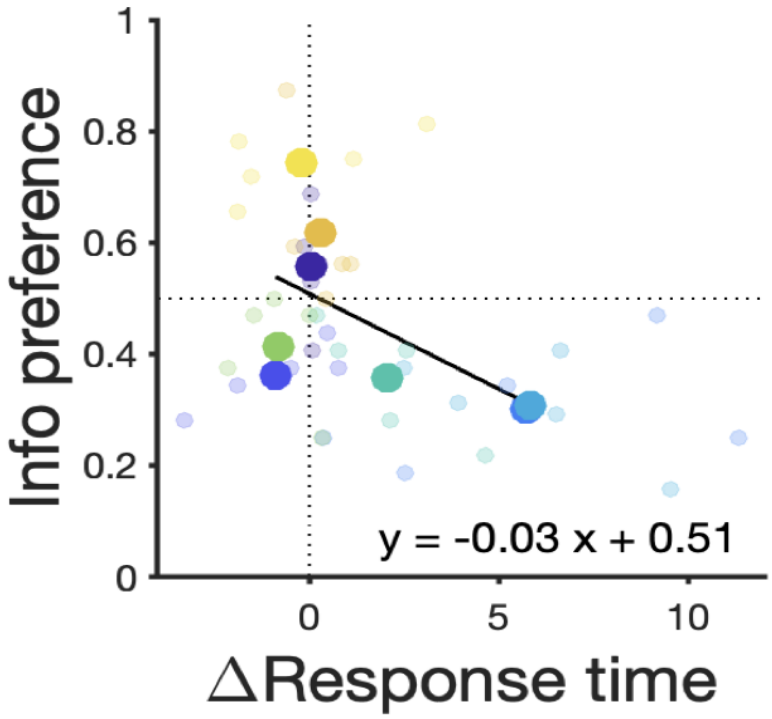


Figure S2. **Preference for the informative option (Info) is correlated with response times.** Mean proportion of choices from the last 5 days of experiment 1 as a function of the mean difference (delta) between Info and NoInfo response times in 1-option forced trials (large markers). Smaller makers depict individual means across sessions, with each colour representing a different individual. Solid black line shows a Deming regression (Estimated model: y = −0.03x + 0.51; slope 95% CI = [-0.07 to 0.33]). n=8.

Author List: Victor Ajuwon*, Tiago Monteiro*, Mark Walton and Alex Kacelnik

Author contributions: VA, TM MW and AK, conceptualised and designed the experiment; VA and TM collected the data; VA, and TM and AK analysed the data; VA and TM wrote the first draft of the paper; all authors ae currently reviewing the manuscript. *Both authors contributed equally to this work.

General Discussion

I this thesis, I set out to investigate an intriguing behavioural phenomenon: in the laboratory individuals from a range of mammal and bird species, including humans, show robust learned preferences for information that has no instrumental use, and even incur significant reward losses in the process. This demand for information that cannot be used to adaptively modify behaviour appears paradoxical with respect to classical normative (Mas-Colell et al., 1995; Stephens and Krebs, 1986; Vasconcelos et al., 2017) and mechanistic (Rescorla and Wagner, 1972) theories of decision-making contenting that individuals are rational agents aimed at maximising well-defined external commodities that ultimately translate into fitness benefits (see Kacelnik, 2006). Instead, these empirical observations suggest that independent of an ability to improve goal-directed outcomes, informative signals themselves are valued by decision-makers.

Over many years the question of whether information *per se* can drive behaviour has been addressed independently in a number of fields including economics (Caplin and Leahy, 2001; Kreps and Porteus, 1978)), experimental and cognitive psychology (Berlyne, 1966, 1954; Loewenstein, 1994; Mackintosh, 1974) and machine learning (Kaplan and Oudeyer, 2007), and has attracted the interest of eminent scientists such as Skinner (1938) and Pavlov (1927). Recently, comparative studies on humans and animals employing the paradoxical choice

paradigm have helped to bridge the gap between these different fields, facilitating rich empirical and theoretical advances aimed at uncovering general principles underlying laboratory observations of non-instrumental information-seeking behaviour across a range of species (Bromberg-Martin and Monosov, 2020; Dunn et al., 2023; Gottlieb and Oudeyer, 2018; Sharot and Sunstein, 2020; Vasconcelos et al., 2018).

In the work presented here, I have used the paradoxical choice task to investigate the mechanisms underlying information value (Chapter 1), how non-instrumental information-seeking affects decision-making in broader contexts (Chapter 2), and the extent to which non-instrumental information-seeking is prevalent across diverse taxa within the vertebrate clade (Chapters 3 and 4). The paradoxical choice task is particularly useful to answer these questions for a number of reasons. First, the task is arbitrarily configured, and thus requires participants to learn the contingencies of the available alternatives, meaning that observed preferences reflect the attribution of value that motivates learning (given that experiments are counterbalanced to control for possible side biases), rather than innate action sequences. Second, the information provided by reward predictive stimuli in the informative option cannot be used by subjects in the task to improve reward gains. Preference for the informative option in paradoxical choice thus represents a particularly interesting case of information-seeking where information has no objective benefit (though see Mackintosh 1974 for the

argument that advanced information allows subjects to prepare for outcomes), and is therefore ideal for studying the subjective value that individuals attribute to information. Third, when the task in implemented such that both options are equally profitable - as in this thesis - both alternatives only differ in their ability to provide information about forthcoming outcomes, avoiding other confounding factors such as differences in reward variance (which can have significant impacts on preference as explored in Chapter 2).

In Chapter 1, I examined the role that two psychological mechanisms play in driving preference for advanced information in paradoxical choice (Ajuwon et al., 2022). One possibility is that reductions in uncertainty in the informative option reinforce choices, while the other is based on conditioned reinforcement: signals for good news (imminent reward), specifically, acquire reinforcing properties while signals for bad news (little/no reward) acquire weak inhibitory properties. To disambiguate these proposals, I investigated the relative effect of both stimuli in the informative option by omitting the salient signal for sure reward and the salient signal for no reward respectively. The results from Chapter 1 were not entirely consistent with strong exclusive support for either hypothesis in question, and thus are open to two possible interpretations.

Across three groups, I found that rat preferences as expressed by choice proportions at the end of the experiment were broadly robust to stimulus

manipulations, with subjects virtually exclusively selecting the informative alternative when both salient stimuli were present, or just when either one was. The fact that subjects preferred the informative alternative when only a salient 'bad news' stimulus was present supports the view that reductions in uncertainty drive preference rather than value attributed to signals for good news specifically. However, the speed of preference acquisition differed between the groups in a manner unpredicted by the uncertainty reduction hypothesis but consistent with the conditioned reinforcement account: omitting a salient stimulus for good news slowed acquisition, while omitting a salient stimulus for bad news weakly enhanced acquisition. Based on these results, it is parsimonious to conclude that while uncertainty reduction is sufficient to generate information-seeking preferences, conditioning of stimuli based on their valence simultaneously acts to modulate the acquisition of preference, a notion consistent with results from macaque (Daddaoua et al., 2016) and human (Kobayashi et al., 2019) studies.

Another possibility is that in the group where there was only a bad news stimulus in the informative option, subjects preferred the informative option because they learned to consider the absence of a salient good news stimulus as good news itself , rather than uncertainty reduction from bad news, meaning that conditioned reinforcement alone could ultimately accounts for informative preferences. This would require further assumptions to explain differences in acquisition speed between the groups, but at present is a possibility. Further work applying

reinforcement learning models to simulate preference acquisition using a similar experimental design may help to clarify the underlying mechanisms. To summarise, the results from Chapter 1 highlight that (1) preference for advanced non-instrumental information is modulated by the salience of reward predictive signals in a manner consistent with a reinforcing effect of salient good news stimuli and a weak inhibitory effect of bad news stimuli (conditioned reinforcement) (2) a salient bad news stimulus alone is sufficient to generate preference in the informative option, indicating that uncertainty reduction may drive informative preferences alongside conditioning effects.

The work in Chapter 1 aimed to contribute to a debate that has been developing for decades, within the animal psychology literature concerning the way in which reward predictive stimuli motivate learning and behaviour. In Chapter 1, though conditioned reinforcement (the action of reward predictive stimuli that have acquired reinforcing properties secondarily via association with primary reinforcers) is presented as one important mechanistic hypothesis to explain subjective value attribution in the context of paradoxical choice, conditioned reinforcement comprises a range of models and hypotheses describing how reward predictive stimuli reinforce behaviour, and has a history the pre-dates the contemporary paradoxical choice protocol. One influential proposal made by Fantino decades ago, Delay Reduction Theory (DRT) proposes that events (i.e., signals) acquire secondary reinforcing properties to the extent that they signal a

reduction in the average waiting time to food (Fatino, 1969,1977). This theory has been extended and revised for application to the paradoxical choice task, inspiring the Signals for Good News Hypothesis (SiGN hypothesis; Dunn and Spetch, 1990; Dunn et al. 2023; McDevitt et al. 2016). One important distinction between DRT and the SiGN hypothesis is that the SiGN hypothesis argues that reward predictive stimuli acquire reinforcing properties by signalling a reduction in waiting time to food relative to the choice response, rather than the average delay to food, since the choice response itself in paradoxical choice is informative regarding time to reward. Another important distinction is that with the SiGN hypothesis signals for bad news do not play a role, in other words they do not acquire inhibitory properties by signalling that the delay to reward will be longer than average relative to the choice response. This asymmetry in the effects of signals for good and bad news, where the potential effects of signals for bad news on choice are presumed to be downweighed or non-existent is characteristic of other conditioned reinforcement hypotheses such as the Predictive (Stimulus) Value hypothesis (Zentall, 2016) and Cunningham and Shahan's (2018) temporal information-theoretic approach.

The assumed lack of an effect of signals for bad news is a key distinguishing feature of conditioned reinforcement hypotheses when compared to a more explicitly informational account of paradoxical choice which we referred to as the ' information hypothesis' in Chapter 1. As with conditioned reinforcement,

there are now a range of explicitly informational hypotheses aimed at explaining paradoxical choice, but a key component of these hypotheses is that stimuli reinforce behaviour because reductions in uncertainty have primary reinforcing value, and these hypotheses predict that the valence of the stimuli themselves should have little effect on choice. This was first formalised by Hendry (1969), who quantified uncertainty reduction in terms of Shannon (1948) entropy, though more recent approaches emphasise the ability of reward predictive stimuli to reduce temporally prolonged uncertainty (Bennett et al., 2016), and it is still debated whether uncertainty reduction acts principally as positive or negative reinforcement. One benefit of explicitly informational approaches is that they focus on a commodity – uncertainty reduction – which has important functional implications for individuals in natural environments (Dall et al., 2005), and therefore aid in the development of models that combine both mechanistic and functional considerations (e.g., Vasconcelos et al., 2015). On the other hand one weakness of this approach is that there can be ambiguity as to the specific definition of uncertainty reduction, though Dall et al.'s (2015) quantitative Bayesian approach may be instructive for future theoretical developments.

Recently, Cunningham and Shahan (2018) developed the temporal information-theoretic model which incorporates approaches used in previous information-based models, while still assuming that conditioned reinforcement is the key driver of choice. According to this account good news stimuli acquire secondary

reinforcing properties by reducing uncertainty about when food will arrive (quantified using Shannon entropy), while stimuli for bad news have no effect on choice. The model also incorporates option profitability and predictive stimulus value as important factors, highlighting that in order to accurately predict behaviour, future models may benefit from several interacting mechanistic factors being considered jointly, which may well include conditioned reinforcement effects alongside valence independent uncertainty reduction.

Having demonstrated robust preferences for non-instrumental information in rats in Chapter 1, in Chapter 2 I aimed to examine the extent to which these preferences can influence decision-making in broader contexts. Previous results have shown that preferences for advanced information can supersede rate-maximisation, raising the question of how else information value might influence decision-making. To do this, I explored how preferences for advanced information influence decision-making under risk, an instance where like in paradoxical choice, individuals display consistent preferences between alternatives yielding the same average reward rate, but differing in another salient dimension: variability in amount of reward.

Previous studies on a range of species have found that individuals are generally risk averse for amounts of reward (Kacelnik and Bateson, 1996): they prefer an alternative consistently offering a fixed intermediate amount of reward over one that unpredictably varies between small and larger amounts. Consistent with this trend and mechanistic theories of risk sensitivity (reviewed in Kacelnik and El Mouden, 2013), I found that rats were generally risk averse for amounts of reward, preferring a fixed alternative over variable one. However, when the variable option provided non-instrumental information in the form of stimuli that eliminated post-choice unpredictability, individuals were risk prone preferring the risk. This switch to risk proneness cannot be readily accounted for by current functional or mechanistic accounts of risk sensitivity, and was presumably driven by the reinforcing value of information. This result highlights that non-instrumental information-seeking can have impacts in a range of decision-making domains, and that classical theories of decision-making should incorporate subjective information value into future analyses. The functional implications of this result are not entirely clear, but one possibility is that in natural foraging situations individuals might use explicit cues within the environment to flexibly modify their responses to risk.

The results from Chapter 2 are also informative regarding the potential mechanisms underlying information value. Chapter 1 highlighted that preferences for non-instrumental information may be driven by both conditioned

reinforcement and uncertainty reduction, but uncertainty reduction could reinforce behaviour in a variety of ways. One possibility is that uncertainty is aversive (Bennett et al., 2016), while another is that the resolution of uncertainty is appetitive (Bromberg-Matin and Hikosaka 2009; Daddaoua et al., 2016). The fact that subjects preferred the informative option with outcomes that are uncertain pre-choice, but made certain post-choice, over the fixed alternative that is associated with no uncertainty, supports the view that subjects seek the resolution of uncertainty, rather than avoiding it altogether. Taken together, the results from Chapters 1 and 2 show that rats have a robust preference for advanced non-instrumental information in the lab, that this is potentially governed by a range of mechanisms, and that this preference can influence decision-making broader contexts, specifically by increasing risk proneness. This highlights rats as a useful model for the study of information value: they may be particularly beneficial in neuroscientific studies of information seeking where macaques are used in very few numbers per study.

While Chapter 1 and Chapter 2 are primarily concerned with Tinbergen's (1963) proximate questions relating to how behaviour is acquired and driven in real time (albeit with functional considerations), Chapter 4 is more concerned with ultimate questions e.g., how widespread is non-instrumental information-seeking within the vertebrate phylogeny? The motivation for tackling this question is that answers will help us better understand the evolution of non-instrumental

information-seeking behaviour and its adaptive significance, while also shedding light on the generality of the underlying mechanisms. In Chapter 4, I carried out the paradoxical choice task in goldfish, a representative of a lineage that is phylogenetically distantly related to those that have so far been tested (mammals and birds). Goldfish are an ideal starting point to explore this question because though relatively under-utilised, they are an established model in cognitive research (Bitterman, 1975; Salena et al., 2021), are strongly motivated by food in instrumental tasks, and their behaviour can be reinforced by conditioned stimuli (Salzinger et al., 1968), making them a good candidate for non-instrumental information-seeking behaviour.

To carry out these experiments with a high degree of experimental refinement, I co-developed *GoFish* (Ajuwon et al., 2023), an automated platform for closed-loop learning experiments on fish and other aquatic species, that is presented in Chapter 3. *GoFish* uses real-time video detection and tracking of subjects to control task contingencies such as the display of visual stimuli on a screen, and the delivery of food rewards using a custom-built pellet dispenser. *GoFish* is operated using Bonsai, a user-friendly, open-source programming language, and the hardware components are cheap and easily accessible. The ability to run experiments with full automation eliminates the need for experimenters to manually deliver rewards and record data, a typical practise in studies of fish cognition that introduces human error and limits scalability. Together, the

refinements afforded by *GoFish*, along with its accessibility and adaptability, make it a useful tool for comparative cognitive research across a range of paradigms. Importantly, we showed that with *GoFish* goldfish learnt to self-initiate trials, and successfully completed a spatial reversal learning experiment and a colour discrimination experiment, validating the procedures employed and highlighting the efficacy of *GoFish* for use in studies of fish learning.

In Chapter 4, I leveraged *GoFish* to conduct the paradoxical choice task on goldfish, enabling a similar level of automation and control found in other investigations of non-instrumental information-seeking. I found that like mammals and birds tested previously (e.g., Ajuwon et al., 2023; Ojeda et al., 2018; Vasconcelos et al., 2015) goldfish learnt the contingencies of the reward predictive stimuli in each option, but unlike other vertebrates so far, goldfish did not show a preference for advanced informative about reward outcomes by choosing the informative option. Subjects underwent extensive training, and task parameters were selected that have been sufficient to generate robust preferences for advanced information in other species, therefore this result likely reflects a genuine species difference in value attributed to non-instrumental information. It is not entirely clear why this species difference exists, but from a mechanistic perspective it highlights that the conditioning of 'good news' stimuli does not necessarily translate into informative preferences. This could be because stimuli for bad news acquire strong inhibitory properties, or because subjects are not

sensitive to reductions in uncertainty. Future work in goldfish should aim to disambiguate these proposals, potentially through stimulus manipulations as in Chapter 1.

From a functional perspective, the difference in preference between goldfish and other vertebrate species shown in Chapter 4, may reflect the fact that fitness benefits acquired from advanced information are dependent on foraging mode. Vasconcelos et al. (2015) argued that preferences for advanced non-instrumental information in the lab may reflect adaptive mechanisms that are rate maximising in natural circumstances, where information about forthcoming rewards will likely be useable. The implication of this perspective is that rather than reflecting subjective information value, preferences in paradoxical choice are generated by mechanisms aimed at maximising reward rate. However, for organisms like carp that are primarily suction feeders and scavengers, such information may not be as ecologically relevant. The result from Chapter 4 therefore raises the possibility that differences in species ecology, in particular foraging mode, may explain disparities in preference for advanced non-instrumental information in the lab. Future work should explore this proposal by using the paradoxical choice task to investigate information-seeking behaviour in predatory fish such as the archerfish (*Toxotes chatareus*), an emerging model in fish cognition (Newport et al., 2016, 2014, 2013).

Further investigations of paradoxical choice in teleost fish species will also help to confirm the absence of non-instrumental information-seeking in this taxon, with implications for our understanding of its evolutionary origins. Testing many different fish species would be difficult, but extending this work to a species like archerfish, that are not domesticated and have a very different ecology to goldfish, will help to address the generality of my preliminary findings. If confirmed, this difference with birds and mammals would suggest that non-instrumental information-seeking behaviour was either acquired by the common ancestor of tetrapods, or evolved in these two taxa by convergence. It would also suggest that a neocortex, or analogous brain structure such as the nidopallium caudolaterale in birds (both linked to other forms of advanced cognition), may be required for explicit computations of uncertainty, consistent with findings in macaques (Blanchard et al., 2015; White et al., 2019).

Looking Forward

Though the use of the paradoxical choice paradigm has advanced our understanding of information-seeking behaviour, studies so far do have limitations. One key issue is that though information in the task is non-instrumental and cannot be used to improve reward gains in the laboratory, the information usually pertains to forthcoming food rewards. This means that it is not entirely clear whether the mechanisms involved ultimately evolved to deal with usable information in a goal-directed, reward maximising manner, or whether the mechanisms reflect subjective value attributed to information itself. As Vasconcelos et al. (2015) argued, such preferences might constitute a misfiring of mechanisms evolved to deal with problems of foraging, rather than reflecting an intrinsic desire to reduce uncertainty, or 'curiosity-like' behaviour aimed at building an accurate world model, as others have suggested (Cervera et al., 2020; Gottlieb and Oudeyer, 2018). Resolving this issue is important, because it has implications for the development of appropriate models in information-seeking and foraging research.

One solution is to adapt the task so that individuals can seek information about other biologically relevant events that have important fitness consequences. One possibility with gregarious species such as rats is to offer advanced information about the possibility of exposure to conspecifics, with 'good news' stimuli

announcing a forthcoming period of interaction with a familiar nest-mate, for example. Starlings have been shown to choose to forage with conspecifics rather than solitarily, even when doing so results in lower rate of gain (Vásquez and Kacelnik, 2000), indicating that animals may be sensitive to non-instrumental information about the potential of such interactions. However, while subjects show tendencies to be in proximity with conspecifics, it is not clear whether this prospect alone can reinforce choice responses in laboratory tasks without a primary reinforcer such as food. Another approach using rats could be to explore individuals' responses to advanced information about whether they will be tickled. Tickling in rats has been shown to induce states of positive affect, and can serve as a reinforcer in operant conditioning (Cloutier et al., 2018; LaFollette et al., 2017), mimicking aspects of playful rough-and-tumble behaviour. This highlights the possibility of using tickling as a reinforcer in future studies of non-instrumental information-seeking, in order to dissociate subjects' responses from forthcoming food rewards, and to distinguish between information-seeking mechanisms more broadly aimed at reducing uncertainty compared to those adapted for specific foraging scenarios. Ultimately, these experimental innovations should help to expose the generality and impact of the theoretical accounts so far proposed.

Another limitation of current research using the paradoxical choice task is that within Tinbergen's (1963) framework, two important aspects of non-instrumental information-seeking behaviour have been relatively neglected: ontogeny and evolutionary history. Studying both of these elements of behaviour comes with challenges: to study ontogeny the contingencies of subjects differing in developmental stages needs to remain similar over time, while studying evolutionary history requires novel behavioural models from unexplored taxa.

The common cuttlefish (*Sepia officinalis*), an emerging model in cognitive research, is ideal for tackling these questions. Cuttlefish are cephalopod molluscs that are strongly motivated by food rewards, show awareness of future events (Billard et al., 2020), and have excellent visual-spatial cognitive skills (Scatà et al., 2016), including episodic-like memories (Jozet-Alves et al., 2013), indicating that they would be able to quickly learn the contingencies of the paradoxical choice task. Furthermore, with minimal modifications *GoFish* could be used to conduct the experiments with full automation and a similar level of control and refinement used with other species so far tested. Cuttlefish diverged from the vertebrate lineage 500 million years ago, and independently evolved sophisticated cognition under different socioecological conditions to traditional vertebrate models (Amodio et al., 2019; Schnell et al., 2021a). Therefore, observing non-instrumental information-seeking in this species would confirm whether this trait has evolved more than once, elucidate its ubiquity across the

whole animal taxon rather than just vertebrates, and reveal whether divergent socio-ecological selection pressures and neural architectures can give rise to non-instrumental information-seeking behaviour. Moreover, cuttlefish have relatively short life-spans (around 2 years), making them an ideal model to investigate the development of traits across the entire life-span (Schnell et al., 2021b). Investigating the ontogeny of paradoxical choice in cuttlefish may also provide functional insights. One hypothesis is that non-instrumental information-seeking behaviour functions to enable younger individuals to acquire knowledge that is useful for solving novel problems in the future (Kidd and Hayden, 2015). Examining developmental changes could corroborate this proposal if younger individuals show a stronger preference for information in the paradoxical choice task information than older, more experienced individuals. For these reasons, leveraging novel experimental technology (Ajuwon et al., 2023) to explore paradoxical choice in cuttlefish promises to address key gaps in our understanding about the development and evolution of non-instrumental information-seeking behaviour.

To conclude, the work in this thesis has exposed the extent to which purely reward-maximising theorizing is insufficient to account for learned behaviour, highlighting that individuals likely ascribe subjective value to information for a variety of reasons, with potentially far-reaching implications on decision-making. The ubiquity of these paradoxical preferences across the animal kingdom is yet to be determined, but the comparative data and novel experimental tools presented in this thesis should help a clearer picture emerge. The future looks bright for this exciting field, that will continue to benefit from a rich, interdisciplinary interchange between theoretical and empirical advances, and the integration of functional and mechanistic accounts of behaviour.

References

Ajuwon, V., Cruz, B.F., Carriço, P., Kacelnik, A., Monteiro, T., Champalimaud Research Scientific Hardware Platform, 2023. GoFish: A low-cost, open-source platform for closed-loop behavioural experiments on fish. Behav Res. https://doi.org/10.3758/s13428-022-02049-2

Ajuwon, V., Ojeda, A., Murphy, R.A., Monteiro, T., Kacelnik, A., 2022. Paradoxical choice and the reinforcing value of information. Animal Cognition. https://doi.org/10.1007/s10071-022-01698-2

Akam, T., Lustig, A., Rowland, J. M., Kapanaiah, S. K., Esteve-Agraz, J., Panniello, M., Márquez, C., Kohl, M. M., Kätzel, D., Costa, R. M., & Walton, M. E. (2022). Open-source, Python-based, hardware and software for controlling behavioural neuroscience experiments. eLife, 11. https://doi.org/10.7554/eLife.67846

Alba, R., Rodríguez, W., Martínez, M., Orduña, V., 2018. Rats' preferences in the suboptimal choice procedure: Evaluating the impact of reinforcement probability and conditioned inhibitors. Behav. Processes 157, 574–582. https://doi.org/10.1016/j.beproc.2018.04.013

Amodio, P., Boeckle, M., Schnell, A.K., Ostojíc, L., Fiorito, G., Clayton, N.S., 2019. Grow Smart and Die Young: Why Did Cephalopods Evolve Intelligence? Trends in Ecology & Evolution 34, 45–56. https://doi.org/10.1016/j.tree.2018.10.010

Aoki, R., Tsuboi, T., & Okamoto, H. (2015). Y-maze avoidance: an automated and rapid associative learning paradigm in zebrafish. Neuroscience Research, 91, 69–72.

Arce, W., & Stevens, J. R. (2022). A precise dispenser design for canine cognition research. Journal of Open Hardware, 6(1). https://doi.org/10.5334/joh.41

Aw, J. M., Holbrook, R. I., Burt de Perera, T., & Kacelnik, A. (2009). State-dependent valuation learning in fish: banded tetras prefer stimuli associated with greater past deprivation. Behavioural Processes, 81(2), 333–336.

Barrus, M.M., Winstanley, C.A., 2016. Dopamine D3 Receptors Modulate the Ability of Win-Paired Cues to Increase Risky Choice in a Rat Gambling Task. J. Neurosci. 36, 785–794. https://doi.org/10.1523/JNEUROSCI.2225-15.2016

Bateson, M., Kacelnik, A., 1995. Preferences for fixed and variable food sources: Variability in amount and delay. Journal of the Experimental Analysis of Behavior 63, 313–329. https://doi.org/10.1901/jeab.1995.63-313

Behrens, T.E.J., Woolrich, M.W., Walton, M.E., Rushworth, M.F.S., 2007. Learning the value of information in an uncertain world. Nat. Neurosci. 10, 1214–1221. https://doi.org/10.1038/nn1954

Beierholm, U.R., Dayan, P., 2010. Pavlovian-Instrumental Interaction in 'Observing Behavior.' PLOS Computational Biology 6, e1000903. https://doi.org/10.1371/journal.pcbi.1000903

Bennett, D., Bode, S., Brydevall, M., Warren, H., Murawski, C., 2016. Intrinsic Valuation of Information in Decision Making under Uncertainty. PLoS Comput. Biol. 12, e1005020. https://doi.org/10.1371/journal.pcbi.1005020

Berger-Tal, O., Nathan, J., Meron, E., Saltz, D., 2014. The Exploration-Exploitation Dilemma: A Multidisciplinary Framework. PLOS ONE 9, e95693. https://doi.org/10.1371/journal.pone.0095693

Berlyne, D.E., 1954. A Theory of Human Curiosity. British Journal of Psychology. General Section 45, 180–191. https://doi.org/10.1111/j.2044-8295.1954.tb01243.x

Berlyne, D.E., 1957. Uncertainty and conflict: A point of contact between information-theory and behavior-theory concepts. Psychol. Rev. 64, 329–339. https://doi.org/10.1037/h0041135

Berlyne, D.E., 1960. Conflict, arousal, and curiosity, Conflict, arousal, and curiosity. McGraw-Hill Book Company, New York, NY, US. https://doi.org/10.1037/11164-000

Berlyne, D.E., 1966. Curiosity and Exploration. Science 153, 25–33. https://doi.org/10.1126/science.153.3731.25

Billard, P., Schnell, A.K., Clayton, N.S., Jozet-Alves, C., 2020. Cuttlefish show flexible and future-dependent foraging cognition. Biology Letters 16, 20190743. https://doi.org/10.1098/rsbl.2019.0743

Bishop, M., Weinhold, M., Turk, A. Z., Adeck, A., & SheikhBahaei, S. (2022). An open-source tool for automated analysis of breathing behaviors in common marmosets and rodents. eLife, 11. https://doi.org/10.7554/eLife.71647

Bitterman, M.E., 1975. The Comparative Analysis of Learning. Science 188, 699–709. https://doi.org/10.1126/science.188.4189.699

Bitterman, M. E. (2006). Classical Conditioning Since Pavlov. Review of General Psychology. https://doi.org/10.1037/e504242006-005

Blanchard, T.C., Hayden, B.Y., Bromberg-Martin, E.S., 2015. Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. Neuron 85, 602–614. https://doi.org/10.1016/j.neuron.2014.12.050

Blaser, R. E., & Vira, D. G. (2014). Experiments on learning in zebrafish (Danio rerio): a promising model of neurocognitive function. Neuroscience and Biobehavioral Reviews, 42, 224–231.

Bower, G., McLean, J., Meacham, J., 1966. Value of knowing when reinforcement is due. J. Comp. Physiol. Psychol. 62, 184–192. https://doi.org/10.1037/h0023682

Brito e Abreu, F., Kacelnik, A., 1999. Energy budgets and risk-sensitive foraging in starlings. Behavioral Ecology 10, 338–345. https://doi.org/10.1093/beheco/10.3.338

Brock, A. J., Sudwarts, A., Daggett, J., Parker, M. O., & Brennan, C. H. (2017). A fully automated computer-based "Skinner Box" for testing learning and memory in zebrafish. bioRxiv. https://doi.org/10.1101/110478

Bromberg-Martin, E.S., Hikosaka, O., 2009. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. Neuron 63, 119–126. https://doi.org/10.1016/j.neuron.2009.06.009

Bromberg-Martin, E.S., Hikosaka, O., 2011. Lateral habenula neurons signal errors in the prediction of reward information. Nature Neuroscience 14, 1209–1216. https://doi.org/10.1038/nn.2902

Bromberg-Martin, E.S., Monosov, I.E., 2020. Neural circuitry of information seeking. Current Opinion in Behavioral Sciences, Curiosity (Explore vs Exploit) 35, 62–70. https://doi.org/10.1016/j.cobeha.2020.07.006

Brydevall, M., Bennett, D., Murawski, C., Bode, S., 2018. The neural encoding of information prediction errors during non-instrumental information seeking. Scientific Reports 8, 6134. https://doi.org/10.1038/s41598-018-24566-x

Burt de Perera, T., & Holbrook, R. I., 2012. Three-dimensional spatial representation in freely swimming fish. Cognitive Processing, 13 Suppl 1, S107–S111.

Buscher, N., Ojeda, A., Francoeur, M., Hulyalkar, S., Claros, C., Tang, T., Terry, A., Gupta, A., Fakhraei, L., & Ramanathan, D. S., 2020. Open-source raspberry Pi-based operant box for translational behavioral testing in rodents. Journal of Neuroscience Methods, 342, 108761.

Calhoun, A.J., Chalasani, S.H., Sharpee, T.O., 2014. Maximally informative foraging by Caenorhabditis elegans. eLife 3, e04220. https://doi.org/10.7554/eLife.04220

Caplin, A., Leahy, J., 2001. Psychological Expected Utility Theory and Anticipatory Feelings*. The Quarterly Journal of Economics 116, 55–79. https://doi.org/10.1162/003355301556347

Caraco, T., Blanckenhorn, W.U., Gregory, G.M., Newman, J.A., Recer, G.M., Zwicker, S.M., 1990. Risk-sensitivity: ambient temperature affects foraging choice. Animal Behaviour 39, 338–345. https://doi.org/10.1016/S0003-3472(05)80879-6

Caraco, T., Martindale, S., Whittam, T.S., 1980. An empirical demonstration of risk-sensitive foraging preferences. Animal Behaviour 28, 820–830. https://doi.org/10.1016/S0003-3472(80)80142-4

Case, J.P., Zentall, T.R., 2018. Suboptimal choice in pigeons: Does the predictive value of the conditioned reinforcer alone determine choice? Behav. Processes 157, 320–326. https://doi.org/10.1016/j.beproc.2018.07.018

Castro, D.C., Berridge, K.C., 2014. Advances in the neurobiological bases for food 'liking' versus 'wanting.' Physiology & Behavior, SSIB 2014 136, 22–30. https://doi.org/10.1016/j.physbeh.2014.05.022

Cervera, R.L., Wang, M.Z., Hayden, B.Y., 2020. Systems neuroscience of curiosity. Current Opinion in Behavioral Sciences, Curiosity (Explore vs Exploit) 35, 48–55. https://doi.org/10.1016/j.cobeha.2020.06.011

Charnov, E.L., 1976. Optimal foraging, the marginal value theorem. Theoretical Population Biology 9, 129–136. https://doi.org/10.1016/0040-5809(76)90040-X

Charpentier, C.J., Bromberg-Martin, E.S., Sharot, T., 2018. Valuation of knowledge and ignorance in mesolimbic reward circuitry. PNAS Proceedings of the National Academy of Sciences of the United States of America 115, E7255–E7264. https://doi.org/10.1073/pnas.1800547115

Chase, A. R., & Hill, W., 1999. Reliable operant apparatus for fish: audio stimulus generator, response button, and pellet-dispensing nipple. Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc, 31(3), 470–478.

Chen, D., Zhang, Qing, Tang, W., Huang, Z., Wang, G., Wang, Yongjun, Shi, J., Xu, H., Lin, L., Li, Z., Chi, W., Huang, L., Xia, J., Zhang, X., Guo, L., Wang, Yuanyuan, Ma, P., Tang, J., Zhou, G., Liu, M., Liu, F., Hua, X., Wang, B., Shen, Q., Jiang, Q., Lin, J., Chen, X., Wang, H., Dou, M., Liu, L., Pan, H., Qi, Y., Wu, B., Fang, J., Zhou, Y., Cen, W., He, W., Zhang, Qiujin, Xue, T., Lin, G., Zhang, W., Liu, Z., Qu, L., Wang, A., Ye, Q., Chen, J., Zhang, Y., Ming, R., Van Montagu, M., Tang, H., Van de Peer, Y., Chen, Y., Zhang, J., 2020. The evolutionary

origin and domestication history of goldfish (Carassius auratus). Proceedings of the National Academy of Sciences 117, 29775–29785. https://doi.org/10.1073/pnas.2005545117

Chow, J.J., Smith, A.P., Wilson, A.G., Zentall, T.R., Beckmann, J.S., 2017. Suboptimal choice in rats: Incentive salience attribution promotes maladaptive decision-making. Behavioural Brain Research 320, 244–254. https://doi.org/10.1016/j.bbr.2016.12.013

Cloutier, S., LaFollette, M.R., Gaskill, B.N., Panksepp, J., Newberry, R.C., 2018. Tickling, a Technique for Inducing Positive Affect When Handling Rats. J Vis Exp 57190. https://doi.org/10.3791/57190

Cohen, J.D., McClure, S.M., Yu, A.J., 2007. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philosophical Transactions of the Royal Society B: Biological Sciences 362, 933–942. https://doi.org/10.1098/rstb.2007.2098

Craft, B.B., 2016. Risk-sensitive foraging: changes in choice due to reward quality and delay. Animal Behaviour 111, 41–47. https://doi.org/10.1016/j.anbehav.2015.09.030

Crowell, C.R., Bernhardt, T.P., 1979. The feature-positive effect and sign-tracking behavior during discrimination learning in the rat. Anim. Learn. Behav. 7, 313–317. https://doi.org/10.3758/BF03209675

Cunningham, P.J., Shahan, T.A., 2018. Suboptimal choice, reward-predictive signals, and temporal information. J Exp Psychol Anim Learn Cogn 44, 1–22. https://doi.org/10.1037/xan0000160

Cunningham, P.J., Shahan, T.A., 2019. Rats engage in suboptimal choice when the delay to food is sufficiently long. J Exp Psychol Anim Learn Cogn 45, 301–310. https://doi.org/10.1037/xan0000211

Daddaoua, N., Lopes, M., Gottlieb, J., 2016. Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates. Sci. Rep. 6, 20202. https://doi.org/10.1038/srep20202

Dall, S.R.X., Giraldeau, L.-A., Olsson, O., McNamara, J.M., Stephens, D.W., 2005. Information and its use by animals in evolutionary ecology. Trends in Ecology & Evolution 20, 187–193. https://doi.org/10.1016/j.tree.2005.01.010

Devarakonda, K., Nguyen, K. P., & Kravitz, A. V., 2016. ROBucket: A low cost operant chamber based on the Arduino microcontroller. Behavior Research Methods, 48(2), 503–509.

Dinsmoor, J.A., 1983. Observing and conditioned reinforcement. Behav. Brain Sci. 6, 693–704. https://doi.org/10.1017/S0140525X00017969

Dinsmoor, J.A., Browne, M.P., Lawrence, C.E., 1972. A test of the negative discriminative stimulus as a reinforcer of observing. J. Exp. Anal. Behav. 18, 79–85. https://doi.org/10.1901/jeab.1972.18-79

Dubey, R., Griffiths, T.L., 2020. Reconciling novelty and complexity through a rational analysis of curiosity. Psychol. Rev. 127, 455–476. https://doi.org/10.1037/rev0000175

Dunn, R., Spetch, M.L., 1990. Choice with uncertain outcomes: conditioned reinforcement effects. J Exp Anal Behav 53, 201–218. https://doi.org/10.1901/jeab.1990.53-201

Dunn, R.M., Pisklak, J.M., McDevitt, M.A., Spetch, M.L., 2023. Suboptimal choice: A review and quantification of the signal for good news (SiGN) model. Psychological Review Advance online publication. https://doi.org/10.1037/rev0000416

Engelhardt, F., Woodard, W. T., & Bitterman, M. E. 1973. Discrimination reversal in the goldfish as a function of training conditions. Journal of Comparative and Physiological Psychology, 85(1), 144–150.

Fantino, E. (1977). Conditioned reinforcement: Choice and information. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 313–339). Englewood Cliffs, NJ: Prentice-Hall.

Fantino, E. (1969). Choice and rate of reinforcement. Journal of the Experimental Analysis of Behavior, 12(5), 723–730. doi:10.1901/jeab.1969.12-723

Fantino, E., Silberberg, A., 2010. Revisiting the Role of Bad News in Maintaining Human Observing Behavior. J. Exp. Anal. Behav. 93, 157–170. https://doi.org/10.1901/jeab.2010.93-157

Foley, H.J., Bates, M., 2019. Sensation and Perception. Routledge.

Foley, N.C., Kelly, S.P., Mhatre, H., Lopes, M., Gottlieb, J., 2017. Parietal neurons encode expected gains in instrumental information. Proc. Natl. Acad. Sci. 114, E3315–E3323. https://doi.org/10.1073/pnas.1613844114

Fortes, I., Machado, A., Vasconcelos, M., 2017. Do pigeons (Columba livia) use information about the absence of food appropriately? A further look into suboptimal choice. J. Comp. Psychol. 131, 277–289. https://doi.org/10.1037/com0000079

Fortes, I., Vasconcelos, M., Machado, A., 2016. Testing the boundaries of "paradoxical" predictions: Pigeons do disregard bad news. J. Exp. Psychol. Anim. Learn. Cogn. 42, 336–346. https://doi.org/10.1037/xan0000114

Freidin, E., Kacelnik, A., 2011. Rational Choice, Context Dependence, and the Value of Information in European Starlings (Sturnus vulgaris). Science 334, 1000–1002. https://doi.org/10.1126/science.1209626

Friston K, Schwartenbeck P, Fitzgerald T, Moutoussis M, Behrens T, Dolan RJ. The anatomy of choice: active inference and agency. Front Hum Neurosci. 2013; 7.

Gallistel, C. R. & Gibbon, J., 2000. Time, rate, and conditioning. Psychological Review 107, 289–344. https://doi.org/10.1037/0033-295X.107.2.289

Gatto, E., Lucon-Xiccato, T., Bisazza, A., Manabe, K., & Dadda, M. 2020. The devil is in the detail: Zebrafish learn to discriminate visual stimuli only if salient. Behavioural Processes, 179, 104215.

Gatto, E., Santacà, M., Verza, I., Dadda, M., & Bisazza, A. 2021. Automated Operant Conditioning Devices for Fish. Do They Work? Animals : An Open Access Journal from MDPI, 11(5). https://doi.org/10.3390/ani11051397

Gatto, E., Testolin, A., Bisazza, A., Zorzi, M., & Lucon-Xiccato, T., 2020. Poor numerical performance of guppies tested in a Skinner box. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-73851-1

Geissmann, Q., Garcia Rodriguez, L., Beckwith, E. J., French, A. S., Jamasb, A. R., & Gilestro, G. F., 2017. Ethoscopes: An open platform for high-throughput ethomics. PLoS Biology, 15(10), e2003026.

Gerlai, R., 2014. Fish in behavior research: unique tools with a great promise! Journal of Neuroscience Methods, 234, 54–58.

Gerlai, R., 2017. Zebrafish and relational memory: Could a simple fish be useful for the analysis of biological mechanisms of complex vertebrate learning? Behavioural Processes, 141, 242–250.

Gerlai, R., 2020. Evolutionary conservation, translational relevance and cognitive function: The future of zebrafish in behavioral neuroscience. Neuroscience and Biobehavioral Reviews, 116, 426–435.

Gipson, C.D., Alessandri, J.J.D., Miller, H.C., Zentall, T.R., 2009. Preference for 50% reinforcement over 75% reinforcement by pigeons. Learn. Behav. 37, 289–298. https://doi.org/10.3758/LB.37.4.289

González, V.V., Blaisdell, A.P., 2021. The role of inhibition in the suboptimal choice task. J. Exp. Psychol. Anim. Learn. Cogn. 47, 429–444. https://doi.org/10.1037/xan0000300

González, V.V., Macías, A., Machado, A., Vasconcelos, M., 2020. The Δ–∑ hypothesis: How contrast and reinforcement rate combine to generate suboptimal choice. Journal of the Experimental Analysis of Behavior 113, 591–608. https://doi.org/10.1002/jeab.595

Gottlieb, J., Hayhoe, M., Hikosaka, O., Rangel, A., 2014. Attention, reward, and information seeking. J. Neurosci. 34, 15497–15504. https://doi.org/10.1523/JNEUROSCI.3270-14.2014

Gottlieb, J., Oudeyer, P.-Y., 2018. Towards a neuroscience of active sampling and curiosity. Nature Reviews Neuroscience 19, 758–770. https://doi.org/10.1038/s41583-018-0078-0

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., Baranes, A., 2013. Information-seeking, curiosity, and attention: computational and neural mechanisms. Trends in Cognitive Sciences 17, 585–593. https://doi.org/10.1016/j.tics.2013.09.001

Grafen, A., 2009. Formalizing Darwinism and inclusive fitness theory. Philosophical Transactions of the Royal Society B: Biological Sciences 364, 3135–3141. https://doi.org/10.1098/rstb.2009.0056

Grafen, A., Hails, R., 2002. Modern Statistics for the Life Sciences. Oxford University Press, Oxford, New York.

Gravelle, H., Rees, R., 2004. Microeconomics. Pearson Education.

Guidelines for the treatment of animals in behavioural research and teaching., 2006. Animal Behaviour, 71(1), 245–253.

Guilbeault, N. C., Guerguiev, J., Martin, M., Tate, I., & Thiele, T. R. (2021). BonZeb: open-source, modular software tools for high-resolution zebrafish tracking and analysis. Scientific Reports, 11(1), 8148.

Gurley, K., 2019. Two open source designs for a low-cost operant chamber using Raspberry Pi™. In Journal of the Experimental Analysis of Behavior (Vol. 111, Issue 3, pp. 508–518). https://doi.org/10.1002/jeab.520

Hendry, D.P., 1969. Conditioned reinforcement, Conditioned reinforcement. Dorsey, Oxford, England.

Iigaya, K., Hauser, T.U., Kurth-Nelson, Z., O'Doherty, J.P., Dayan, P., Dolan, R.J., 2020. The value of what's to come: Neural mechanisms coupling prediction error and the utility of anticipation. Science Advances 6, eaba3828. https://doi.org/10.1126/sciadv.aba3828

Iigaya, K., Story, G.W., Kurth-Nelson, Z., Dolan, R.J., Dayan, P., 2016. The modulation of savouring by prediction error and its effects on choice. eLife 5, e13747. https://doi.org/10.7554/eLife.13747

Jenkins, H.M., Boakes, R.A., 1973. Observing stimulus sources that signal food or no food. J. Exp. Anal. Behav. 20, 197–207. https://doi.org/10.1901/jeab.1973.20-197

Jozet-Alves, C., Bertin, M., Clayton, N.S., 2013. Evidence of episodic-like memory in cuttlefish. Current Biology 23, R1033–R1035. https://doi.org/10.1016/j.cub.2013.10.021

Kacelnik, A., 2003. The evolution of patience, in: Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice. Russell Sage Foundation, New York, NY, US, pp. 115–138.

Kacelnik, A., 2006. Meanings of rationality, in: Hurley, S., Nudds, M. (Eds.), Rational Animals? Oxford University Press, p87-106. https://doi.org/10.1093/acprof:oso/9780198528272.003.0002

Kacelnik, A., 2007. Normative and Descriptive Models of Decision Making: Time Discounting and Risk Sensitivity, in: Ciba Foundation Symposium 208 - Characterizing Human Psychological Adaptations, Novartis Foundation Symposia. pp. 51–70. https://doi.org/10.1002/9780470515372.ch5

Kacelnik, A., Bateson, M., 1996. Risky Theories—The Effects of Variance on Foraging Decisions. Am Zool 36, 402–434. https://doi.org/10.1093/icb/36.4.402

Kacelnik, A., Bateson, M., 1997. Risk-sensitivity: crossroads for theories of decision-making. Trends in Cognitive Sciences 1, 304–309. https://doi.org/10.1016/S1364-6613(97)01093-0

Kacelnik, A., Brito e Abreu, F., 1998. Risky Choice and Weber's Law. Journal of Theoretical Biology 194, 289–298. https://doi.org/10.1006/jtbi.1998.0763

Kacelnik, A., El Mouden, C., 2013. Triumphs and trials of the risk paradigm. Animal Behaviour 86, 1117–1129. https://doi.org/10.1016/j.anbehav.2013.09.034

Kacelnik, A., Vasconcelos, M., Monteiro, T., Aw, J., 2011. Darwin's "tug-of-war" vs. starlings' "horse-racing": how adaptations for sequential encounters drive simultaneous choice. Behav. Ecol. Sociobiol. 65, 547–558. https://doi.org/10.1007/s00265-010-1101-2

Kalueff, A. V., Gebhardt, M., Stewart, A. M., Cachat, J. M., Brimmer, M., Chawla, J. S., Craddock, C., Kyzar, E. J., Roth, A., Landsman, S., Gaikwad, S., Robinson, K., Baatrup, E., Tierney, K., Shamchuk, A., Norton, W., Miller, N., Nicolson, T., Braubach, O., … and the Zebrafish Neuros., 2013. Towards a Comprehensive Catalog of Zebrafish Behavior 1.0 and Beyond. Zebrafish, 10(1), 70–86.

Kalueff, A. V., Stewart, A. M., & Gerlai, R., 2014. Zebrafish as an emerging model for studying complex brain disorders. Trends in Pharmacological Sciences, 35(2), 63–75.

Kane, G. A., Lopes, G., Saunders, J. L., Mathis, A., & Mathis, M. W., 2020. Real-time, low-latency closed-loop feedback using markerless posture tracking. eLife, 9. https://doi.org/10.7554/eLife.61909

Kapanaiah, S. K. T., van der Veen, B., Strahnen, D., Akam, T., & Kätzel, D., 2021. A low-cost open-source 5-choice operant box system optimized for electrophysiology and optophysiology in mice. Scientific Reports, 11(1), 22279.

Kaplan, F., Oudeyer, P.-Y., 2007. In search of the neural circuits of intrinsic motivation. Frontiers in Neuroscience 1.

Kaplan, F., Oudeyer, P.-Y., 2007. Intrinsically Motivated Machines, in: Lungarella, M., Iida, F., Bongard, J., Pfeifer, R. (Eds.), 50 Years of Artificial Intelligence: Essays Dedicated to the 50th Anniversary of Artificial Intelligence, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 303–314. https://doi.org/10.1007/978-3-540-77296-5_27

Kendall, S.B., 1973. Redundant information in an observing-response procedure. J. Exp. Anal. Behav. 19, 81–92. https://doi.org/10.1901/jeab.1973.19-81

Kendall, S.B., 1974. Preference for intermittent reinforcement. J Exp Anal Behav 21, 463–473. https://doi.org/10.1901/jeab.1974.21-463

Kidd, C., Hayden, B.Y., 2015. The psychology and neuroscience of curiosity. Neuron 88, 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Knight, F.H., 1921. Risk, Uncertainty and Profit. Boston and New York, Houghton Mifflin Company.

Kobayashi, K., Hsu, M., 2019. Common neural code for reward and information value. Proc. Natl. Acad. Sci. 116, 13061–13066. https://doi.org/10.1073/pnas.1820145116

Kobayashi, K., Ravaioli, S., Baranès, A., Woodford, M., Gottlieb, J., 2019. Diverse motives for human curiosity. Nat Hum Behav 3, 587–595. https://doi.org/10.1038/s41562-019-0589-3

Krebs, J.R., Kacelnik, A., Taylor, P., 1978. Test of optimal sampling by foraging great tits. Nature 275, 27–31. https://doi.org/10.1038/275027a0

Kreps, D.M., Porteus, E.L., 1978. Temporal Resolution of Uncertainty and Dynamic Choice Theory. Econometrica 46, 185–200. https://doi.org/10.2307/1913656

Kuroda, T., Mizutani, Y., Cançado, C. R. X., & Podlesnik, C. A., 2017. Reversal learning and resurgence of operant behavior in zebrafish (Danio rerio). Behavioural Processes, 142, 79–83.

LaFollette, M.R., O'Haire, M.E., Cloutier, S., Blankenberger, W.B., Gaskill, B.N., 2017. Rat tickling: A systematic review of applications, outcomes, and moderators. PLoS One 12, e0175320. https://doi.org/10.1371/journal.pone.0175320

Laude, J.R., Stagner, J.P., Zentall, T.R., 2014. Suboptimal Choice by Pigeons May Result from the Diminishing Effect of Nonreinforcement. J. Exp. Psychol. Anim. Learn. Cogn. 40, 12–21. https://doi.org/10.1037/xan0000010

Lieberman, D.A., 1972. Secondary reinforcement and information as determinants of observing behavior in monkeys (Macaca mulatta). Learn. Motiv. 3, 341–358. https://doi.org/10.1016/0023-9690(72)90030-6

Lieberman, D.A., Cathro, J.S., Nichol, K., Watson, E., 1997. The Role of S− in Human Observing Behavior: Bad News Is Sometimes Better Than No News. Learn. Motiv. 28, 20–42. https://doi.org/10.1006/lmot.1997.0951

Livingston, S., Bennett, W.L. (Eds.), 2020. The Disinformation Age, in: The Disinformation Age, SSRC Anxieties of Democracy. Cambridge University Press, Cambridge.

Lockard, J.S., 1963. Choice of a warning signal or no warning signal in an unavoidable shock situation. J. Comp. Physiol. Psychol. 56, 526–530. https://doi.org/10.1037/h0041552

Loewenstein, G., 1994. The psychology of curiosity: A review and reinterpretation. Psychological Bulletin 116, 75–98. https://doi.org/10.1037/0033-2909.116.1.75

Loewenstein, G., Molnar, A., 2018. The renaissance of belief-based utility in economics. Nat Hum Behav 2, 166–167. https://doi.org/10.1038/s41562-018-0301-z

Lopes, G., & Monteiro, P. 2021. New Open-Source Tools: Using Bonsai for Behavioral Tracking and Closed-Loop Experiments. Frontiers in Behavioral Neuroscience, 15, 647640.

Lopes, G., Bonacchi, N., Frazão, J., Neto, J. P., Atallah, B. V., Soares, S., Moreira, L., Matias, S., Itskov, P. M., Correia, P. A., Medina, R. E., Calcaterra, L., Dreosti, E., Paton, J. J., & Kampff, A. R. 2015. Bonsai: an event-based framework for processing and controlling data streams. Frontiers in Neuroinformatics, 9, 7.

Lopes, G., Farrell, K., Horrocks, E. A., Lee, C.-Y., Morimoto, M. M., Muzzu, T., Papanikolaou, A., Rodrigues, F. R., Wheatcroft, T., Zucca, S., Solomon, S. G., & Saleem, A. B., 2021. Creating and controlling visual environments using BonVision. eLife, 10. https://doi.org/10.7554/eLife.65541

Lopes, G., Monteiro, P., 2021., New Open-Source Tools: Using Bonsai for Behavioral Tracking and Closed-Loop Experiments. Frontiers in Behavioral Neuroscience 15.

López, P., Alba, R., Orduña, V., 2018. Individual differences in incentive salience attribution are not related to suboptimal choice in rats. Behav. Brain Res. 341, 71–78. https://doi.org/10.1016/j.bbr.2017.12.028

Lucon-Xiccato, T., Manabe, K., & Bisazza, A., 2018. Guppies learn faster to discriminate between red and yellow than between two shapes. Ethology. https://doi.org/10.1111/eth.12829

MacArthur, R.H., Pianka, E.R., 1966. On Optimal Use of a Patchy Environment. The American Naturalist 100, 603–609. https://doi.org/10.1086/282454

Macías, A., González, V.V., Machado, A., Vasconcelos, M., 2021. The functional equivalence of two variants of the suboptimal choice task: choice proportion and response latency as measures of value. Anim. Cogn. 24, 85–98. https://doi.org/10.1007/s10071-020-01418-8

MacKay, D.J., 2003. Information theory, inference and learning algorithms. Cambridge university press.

Mackintosh, N.J., 1974. The psychology of animal learning, The psychology of animal learning. Academic Press, Oxford, England.

Manabe, K., Dooling, R. J., & Takaku, S., 2013. An automated device for appetitive conditioning in zebrafish (Danio rerio). Zebrafish, 10(4), 518–523.

Martínez, M., Alba, R., Rodríguez, W., Orduña, V., 2017. Incentive salience attribution is not the sole determinant of suboptimal choice in rats: Conditioned inhibition matters. Behavioural Processes 142, 99–105. https://doi.org/10.1016/j.beproc.2017.06.012

Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic Theory, International edition. ed. OUP USA, New York.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M., 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nature Neuroscience, 21(9), 1281–1289.

Matute, H., Blanco, F., Díaz-Lago, M., 2019. Learning mechanisms underlying accurate and biased contingency judgments. Journal of Experimental Psychology: Animal Learning and Cognition 45, 373–389. https://doi.org/10.1037/xan0000222

McDevitt, M.A., Dunn, R.M., Spetch, M.L., Ludvig, E.A., 2016. When good news leads to bad choices. J. Exp. Anal. Behav. 105, 23–40. https://doi.org/10.1002/jeab.192

McDevitt, M.A., Pisklak, J.M., Dunn, R.M., Spetch, M.L., 2022. Forced-exposure trials increase suboptimal choice. Psychon Bull Rev 29, 1514–1523. https://doi.org/10.3758/s13423-022-02092-2

McDevitt, M.A., Pisklak, J.M., Spetch, M., Dunn, R., 2018. The influence of outcome delay on suboptimal choice. Behav. Processes 157, 279–285. https://doi.org/10.1016/j.beproc.2018.10.008

McDevitt, M.A., Spetch, M.L., Dunn, R., 1997. Contiguity and conditioned reinforcement in probabilistic choice. J. Exp. Anal. Behav. 68, 317–327. https://doi.org/10.1901/jeab.1997.68-317

McNamara, J.M., 1996. Risk-Prone Behaviour under Rules Which Have Evolved in a Changing Environment. American Zoologist 36, 484–495.

McNamara, J.M., Houston, A.I., 1992. Risk-sensitive foraging: A review of the theory. Bulletin of Mathematical Biology 54, 355–378. https://doi.org/10.1016/S0092-8240(05)80031-X

Meshalkina, D. A., Kizlyk, M. N., Kysil, E. V., Collier, A. D., Echevarria, D. J., Abreu, M. S., Barcellos, L. J. G., Song, C., & Kalueff, A. V., 2017. Understanding zebrafish cognition. Behavioural Processes, 141, 229–241.

Miletto Petrazzini, M. E., Pecunioso, A., Dadda, M., & Agrillo, C., 2020. Does brain lateralization affect the performance in binary choice tasks? A study in the animal model Danio rerio. Symmetry, 12(8), 1294.

Monteiro, T., Hart, T., & Kacelnik, A., 2021. Imprinting on time-structured acoustic stimuli in ducklings. Biology Letters, 17(9), 20210381.

Monteiro, T., Vasconcelos, M., & Kacelnik, A., 2020. Choosing fast and simply: Construction of preferences by starlings through parallel option valuation. PLoS Biology, 18(8), e3000841.

Monteiro, T., Vasconcelos, M., Kacelnik, A., 2013. Starlings uphold principles of economic rationality for delay and probability of reward. Proceedings of the Royal Society B: Biological Sciences 280, 20122386. https://doi.org/10.1098/rspb.2012.2386

Monteiro, T., Vasconcelos, M., Kacelnik, A., 2020. Choosing fast and simply: Construction of preferences by starlings through parallel option valuation. PLOS Biol. 18, e3000841. https://doi.org/10.1371/journal.pbio.3000841

Mueller, K. P., & Neuhauss, S. C. F., 2012. Automated visual choice discrimination learning in zebrafish (Danio rerio). Journal of Integrative Neuroscience, 11(1), 73–85.

Murphy, R.A., Baker, A.G. & Fouquet, N., 2001. Relative validity of contextual and discrete cues. Journal of Experimental Psychology: Animal Behavior Processes, 27, 137-152

Neumeyer, C. 1984,. On spectral sensitivity in the goldfish. Evidence for neural interactions between different "cone mechanisms." Vision Research, 24(10), 1223–1231.

Newman, J.P., Wolff, W.T., Hearst, E., 1980. The feature-positive effect in adult human subjects. J. Exp. Psychol. [Hum. Learn.] 6, 630–650. https://doi.org/10.1037/0278-7393.6.5.630

Newport, C., 2021. Abstract concept learning in fish. Current Opinion in Behavioral Sciences, 37, 56–62.

Newport, C., Padget, O., & de Perera, T. B., 2021. High turbidity levels alter coral reef fish movement in a foraging task. In Scientific Reports (Vol. 11, Issue 1). https://doi.org/10.1038/s41598-021-84814-5

Newport, C., Wallis, G., Reshitnyk, Y., & Siebeck, U. E., 2016. Discrimination of human faces by archerfish (Toxotes chatareus). Scientific Reports, 6, 27523.

Newport, C., Wallis, G., Siebeck, U.E., 2014. Concept learning and the use of three common psychophysical paradigms in the archerfish (Toxotes chatareus). Front. Neural Circuits 8. https://doi.org/10.3389/fncir.2014.00039

Newport, C., Wallis, G., Temple, S.E., Siebeck, U.E., 2013. Complex, context-dependent decision strategies of archerfish, Toxotes chatareus. Animal Behaviour 86, 1265–1274. https://doi.org/10.1016/j.anbehav.2013.09.031

O'Leary, J. D., O'Leary, O. F., Cryan, J. F., & Nolan, Y. M., 2018. A low-cost touchscreen operant chamber using a Raspberry Pi^{TM}. Behavior Research Methods, 50(6), 2523–2530.

Oh, J., Hofer, R., & Fitch, T., 2017. An open source automatic feeder for animal experiments. In HardwareX (Vol. 1, pp. 13–21). https://doi.org/10.1016/j.ohx.2016.09.001

Ojeda, A., Murphy, R.A., Kacelnik, A., 2018. Paradoxical choice in rats: Subjective valuation and mechanism of choice. Behav. Processes 152, 73–80. https://doi.org/10.1016/j.beproc.2018.03.024

Orger, M. B., & de Polavieja, G. G. (2017). Zebrafish Behavior: Opportunities and Challenges. Annual Review of Neuroscience, 40, 125–147.

Oudeyer, P.-Y., Kaplan, F., Hafner, V.V., 2007. Intrinsic Motivation Systems for Autonomous Mental Development. IEEE Transactions on Evolutionary Computation 11, 265–286. https://doi.org/10.1109/TEVC.2006.890271

Pace, G.M., McCoy, D.F., Nallan, G.B., 1980. Feature-positive and feature-negative learning in the Rhesus monkey and pigeon. Am. J. Psychol. 93, 409–427. https://doi.org/10.2307/1422721

Parker, G.A., Smith, J.M., 1990. Optimality theory in evolutionary biology. Nature 348, 27–33. https://doi.org/10.1038/348027a0

Pavlov, I.P., 1927. Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. Oxford Univ. Press, Oxford, England.

Pereira, T. D., Tabris, N., Matsliah, A., Turner, D. M., Li, J., Ravindranath, S., Papadoyannis, E. S., Normand, E., Deutsch, D. S., Wang, Z. Y., McKenzie-Smith, G. C., Mitelut, C. C., Castro, M. D., D'Uva, J., Kislin, M., Sanes, D. H., Kocher, S. D., Wang, S. S.-H., Falkner, A. L., … Murthy, M., 2022. SLEAP: A deep learning system for multi-animal pose tracking. Nature Methods, 19(5), 628.

Pineño, O., 2014. ArduiPod Box: A low-cost and open-source Skinner box using an iPod Touch and an Arduino microcontroller. In Behavior Research Methods (Vol. 46, Issue 1, pp. 196–205). https://doi.org/10.3758/s13428-013-0367-5

Pisklak, J.M., McDevitt, M.A., Dunn, R.M., Spetch, M.L., 2015. When good pigeons make bad decisions: Choice with probabilistic delays and outcomes. J. Exp. Anal. Behav. 104, 241–251. https://doi.org/10.1002/jeab.177

Pisklak, J.M., McDevitt, M.A., Dunn, R.M., Spetch, M.L., 2019. Frequency and value both matter in the suboptimal choice procedure. Journal of the Experimental Analysis of Behavior 111, 1–11. https://doi.org/10.1002/jeab.490

Potrich, D., Zanon, M., & Vallortigara, G., 2022. Archerfish number discrimination. eLife, 11. https://doi.org/10.7554/eLife.74057

Pouca, C. V., & Brown, C., 2017. Contemporary topics in fish cognition and behaviour. In Current Opinion in Behavioral Sciences (Vol. 16, pp. 46–52). https://doi.org/10.1016/j.cobeha.2017.03.002

Prokasy Jr., W.F., 1956. The acquisition of observing responses in the absence of differential external reinforcement. J. Comp. Physiol. Psychol. 49, 131–134. https://doi.org/10.1037/h0046740

Rachlin, H., Battalio, R., Kagel, J., Green, L., 1981. Maximization theory in behavioral psychology. Behavioral and Brain Sciences 4, 371–388. https://doi.org/10.1017/S0140525X00009407

Rangel, A., Camerer, C., Montague, P.R., 2008. A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci 9, 545–556. https://doi.org/10.1038/nrn2357

Reboreda, J.C., Kacelnik, A., 1991. Risk sensitivity in starlings: variability in food amount and food delay. Behavioral Ecology 2, 301–308. https://doi.org/10.1093/beheco/2.4.301

Rescorla, R., Wagner, A., 1972. A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. Classical Conditioning: Current Research and Theory.

Rodriguez Cabrero, J.A.M., Zhu, J.-Q., Ludvig, E.A., 2019. Costly curiosity: People pay a price to resolve an uncertain gamble early. Behavioural Processes 160, 20–25. https://doi.org/10.1016/j.beproc.2018.12.015

Roper, K.L., Zentall, T.R., 1999. Observing Behavior in Pigeons: The Effect of Reinforcement Probability and Response Cost Using a Symmetrical Choice Procedure. Learn. Motiv. 30, 201–220. https://doi.org/10.1006/lmot.1999.1030

Rosenström, T., Wiesner, K., Houston, A.I., 2016. Scalar utility theory and proportional processing: What does it actually imply? Journal of Theoretical Biology 404, 222–235. https://doi.org/10.1016/j.jtbi.2016.06.003

Sainsbury, R., 1971. The "feature positive effect" and simultaneous discrimination learning. J. Exp. Child Psychol. 11, 347–356. https://doi.org/10.1016/0022-0965(71)90039-7

Salena, M.G., Turko, A.J., Singh, A., Pathak, A., Hughes, E., Brown, C., Balshine, S., 2021. Understanding fish cognition: A review and appraisal of current practices. Animal Cognition 24, 395–406. https://doi.org/10.1007/s10071-021-01488-2

Salzinger, K., Freimark, S.J., Fairhurst, S.P., Wolkoff, F.D., 1968. Conditioned Reinforcement in the Goldfish. Science 160, 1471–1472. https://doi.org/10.1126/science.160.3835.1471

Santacà, M., Dadda, M., Miletto Petrazzini, M. E., & Bisazza, A., 2021. Stimulus characteristics, learning bias and visual discrimination in zebrafish (Danio rerio). Behavioural Processes, 192, 104499.

Sasaki, T., Pratt, S.C., Kacelnik, A., 2018. Parallel vs. comparative evaluation of alternative options by colonies and individuals of the ant Temnothorax rugatulus. Sci. Rep. 8, 12730. https://doi.org/10.1038/s41598-018-30656-7

Scatà, G., Jozet-Alves, C., Thomasse, C., Josef, N., Shashar, N., 2016. Spatial learning in the cuttlefish Sepia officinalis: preference for vertical over horizontal information. Journal of Experimental Biology 219, 2928–2933. https://doi.org/10.1242/jeb.129080

Schluessel, V., Kreuter, N., Gosemann, I. M., & Schmidt, E., 2022. Cichlids and stingrays can add and subtract "one" in the number space from one to five. Scientific Reports, 12(1), 3894.

Schnell, A.K., Amodio, P., Boeckle, M., Clayton, N.S., 2021a. How intelligent is a cephalopod? Lessons from comparative cognition. Biological Reviews 96, 162–178.

Schnell, A.K., Clayton, N.S., Hanlon, R.T., Jozet-Alves, C., 2021b. Episodic-like memory is preserved with age in cuttlefish. Proceedings of the Royal Society B: Biological Sciences 288, 20211052. https://doi.org/10.1098/rspb.2021.1052

Shahan, T.A., Cunningham, P., 2015. Conditioned reinforcement and information theory reconsidered. J. Exp. Anal. Behav. 103, 405–418. https://doi.org/10.1002/jeab.142

Shannon, C.E., 1948. A Mathematical Theory of Communication. Bell Syst. Tech. J. 27, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shapiro, M.S., Siller, S., Kacelnik, A., 2008. Simultaneous and sequential choice as a function of reward delay and magnitude: Normative, descriptive and process-based models tested in the European starling (Sturnus vulgaris). J. Exp. Psychol. Anim. Behav. Process. 34, 75–93. https://doi.org/10.1037/0097-7403.34.1.75

Sharot, T., Sunstein, C.R., 2020. How people decide what they want to know. Nat Hum Behav 4, 14–19. https://doi.org/10.1038/s41562-019-0793-1

Shettleworth, S. J., 2009. The evolution of comparative cognition: is the snark still a boojum? Behavioural Processes, 80(3), 210–217.

Siegle, J. H., López, A. C., Patel, Y. A., Abramov, K., Ohayon, S., & Voigts, J., 2017. Open Ephys: an open-source, plugin-based platform for multichannel electrophysiology. Journal of Neural Engineering, 14(4), 045003.

Silberberg, A., Fantino, E., 2010. Observing Responses: Maintained by Good News Only? Behav. Processes 85, 80–82. https://doi.org/10.1016/j.beproc.2010.06.002

Skinner, B.F., 1938. The behavior of organisms: an experimental analysis, The behavior of organisms: an experimental analysis. Appleton-Century, Oxford, England.

Slovic, P., 1995. The construction of preference. Am. Psychol. 50, 364–371. https://doi.org/10.1037/0003-066X.50.5.364

Smith, A.P., Bailey, A.R., Chow, J.J., Beckmann, J.S., Zentall, T.R., 2016. Suboptimal Choice in Pigeons: Stimulus Value Predicts Choice over Frequencies. PLOS ONE 11, e0159336. https://doi.org/10.1371/journal.pone.0159336

Smith, A.P., Zentall, T.R., Kacelnik, A., 2018. Midsession reversal task with pigeons: Parallel processing of alternatives explains choices. J. Exp. Psychol. Anim. Learn. Cogn. 44, 272–279. https://doi.org/10.1037/xan0000180

Spetch, M.L., Belke, T.W., Barnet, R.C., Dunn, R., Pierce, W.D., 1990. Suboptimal choice in a percentage-reinforcement procedure: effects of signal condition and terminal-link length. J Exp Anal Behav 53, 219–234. https://doi.org/10.1901/jeab.1990.53-219

Spetch, M.L., Mondloch, M.V., Belke, T.W., Dunn, R., 1994. Determinants of pigeons' choice between certain and probabilistic outcomes. Anim. Learn. Behav. 22, 239–251. https://doi.org/10.3758/BF03209832

Sridhar, V. H., Li, L., Gorbonos, D., Nagy, M., Schell, B. R., Sorochkin, T., Gov, N. S., & Couzin, I. D., 2021. The geometry of decision-making in individuals and collectives. Proceedings of the National Academy of Sciences of the United States of America, 118(50). https://doi.org/10.1073/pnas.2102157118

Staddon, J.E.R., Cerutti, D.T., 2003. Operant Conditioning. Annu. Rev. Psychol. 54, 115–144. https://doi.org/10.1146/annurev.psych.54.101601.145124

Stagner, J.P., Zentall, T.R., 2010. Suboptimal choice behavior by pigeons. Psychon Bull Rev 17, 412–416. https://doi.org/10.3758/PBR.17.3.412

Stephens, D.W., 1981. The logic of risk-sensitive foraging preferences. Animal Behaviour 29, 628–629. https://doi.org/10.1016/S0003-3472(81)80128-5

Stephens, D.W., Krebs, J.R., 1986. Foraging Theory. Princeton University Press.

Stewart, A. M., Grieco, F., Tegelenbosch, R. A. J., Kyzar, E. J., Nguyen, M., Kaluyeva, A., Song, C., Noldus, L. P. J. J., & Kalueff, A. V., 2015. A novel 3D method of locomotor analysis in adult zebrafish: Implications for automated detection of CNS drug-evoked phenotypes. Journal of Neuroscience Methods, 255, 66–74.

Štih, V., Petrucco, L., Kist, A. M., & Portugues, R., 2019. Stytra: An open-source, integrated system for stimulation, tracking and closed-loop behavioral experiments. PLoS Computational Biology, 15(4), e1006699.

Sutton, R.S., Barto, A.G., 2018. Reinforcement learning: An introduction, 2nd ed, Reinforcement learning: An introduction, 2nd ed. The MIT Press, Cambridge, MA, US.

Swanson, K., White, S. R., Preston, M. W., Wilson, J., Mitchell, M., & Laubach, M., 2021. An Open Source Platform for Presenting Dynamic Visual Stimuli. eNeuro, 8(3). https://doi.org/10.1523/ENEURO.0563-20.2021

Talton, L. E., Higa, J. J., & Staddon, J. E., 1999. Interval schedule performance in the goldfish Carassius auratus. Behavioural Processes, 45(1-3), 193–206.

The R Project for Statistical Computing., 2018. https://www.R-project.org

Tinbergen, N., 1963. On aims and methods of Ethology. Zeitschrift für Tierpsychologie 20, 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Trujano, R.E., López, P., Rojas-Leguizamón, M., Orduña, V., 2016. Optimal behavior by rats in a choice task is associated to a persistent conditioned inhibition effect. Behavioural Processes 130, 65–70. https://doi.org/10.1016/j.beproc.2016.07.005

Trujano, R.E., Orduña, V., 2015. Rats are optimal in a choice task in which pigeons are not. Behavioural Processes 119, 22–27. https://doi.org/10.1016/j.beproc.2015.07.010

van Lieshout, L.L., de Lange, F.P., Cools, R., 2020. Why so curious? Quantifying mechanisms of information seeking. Curr. Opin. Behav. Sci., Curiosity (Explore vs Exploit) 35, 112–117. https://doi.org/10.1016/j.cobeha.2020.08.005

Vasconcelos, M., Fortes, I., Kacelnik, A., 2017. On the structure and role of optimality models in the study of behavior, in: APA Handbook of Comparative Psychology: Perception, Learning, and Cognition, Vol. 2, APA Handbooks in Psychology®. American Psychological Association, Washington, DC, US, pp. 287–307. https://doi.org/10.1037/0000012-014

Vasconcelos, M., Machado, A., Pandeirada, J.N.S., 2018. Ultimate explanations and suboptimal choice. Behav. Processes 152, 63–72. https://doi.org/10.1016/j.beproc.2018.03.023

Vasconcelos, M., Monteiro, T., Kacelnik, A., 2015. Irrational choice and the value of information. Sci Rep 5. https://doi.org/10.1038/srep13874

Vásquez, R.A., Kacelnik, A., 2000. Foraging rate versus sociality in the starling Sturnus vulgaris. Proceedings of the Royal Society of London. Series B: Biological Sciences 267, 157–164. https://doi.org/10.1098/rspb.2000.0981

Wallace, K. J., Rausch, R. T., Ramsey, M. E., & Cummings, M. E., 2020. Sex differences in cognitive performance and style across domains in mosquitofish (Gambusia affinis). Animal Cognition, 23(4), 655–669.

Walter, T., & Couzin, I. D., 2021. TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields. eLife, 10. https://doi.org/10.7554/eLife.64000

Wichmann, F.A., Hill, N.J., 2001. The psychometric function: I. Fitting, sampling, and goodness of fit. Percept. Psychophys. 63, 1293–1313. https://doi.org/10.3758/BF03194544

Wyckoff, L.B., 1959. Toward a quantitative theory of secondary reinforcement. Psychol. Rev. 66, 68–78. https://doi.org/10.1037/h0046882

Wyckoff, L.B., 1969. The Role of Observing Responses in Discrimination Learning: Part II. In Conditioned Reinforcement. Ed. D.P. Hendry, 237-260. Homewood, Illinois: The Dorsey Press.

Zentall, T.R., 2013. Animals prefer reinforcement that follows greater effort: Justification of effort or within-trial contrast? Comp. Cogn. Behav. Rev. 8, 60–77. https://doi.org/10.3819/ccbr.2013.80004

Zentall, T.R., 2016. Resolving the paradox of suboptimal choice. J Exp Psychol Anim Learn Cogn 42, 1–14. https://doi.org/10.1037/xan0000085

Zentall, T.R., Laude, J.R., Stagner, J.P., Smith, A.P., 2015. Suboptimal choice by pigeons: Evidence that the value of the conditioned reinforcer rather than its frequency determines choice. The Psychological Record 65, 223–229. https://doi.org/10.1007/s40732-015-0119-2

Zentall, T.R., Smith, A.P., Beckmann, J., 2019. Differences in rats and pigeons suboptimal choice may depend on where those stimuli are in their behavior system. Behavioural Processes 159, 37–41. https://doi.org/10.1016/j.beproc.2018.11.012

Zentall, T.R., Smith, A.P., Beckmann, J.S., 2019. Gambling behavior: An animal model. Translational Issues in Psychological Science 5, 276–288. https://doi.org/10.1037/tps0000200

Zentall, T.R., Stagner, J., 2011. Maladaptive choice behaviour by pigeons: an animal analogue and possible mechanism for gambling (sub-optimal human decision-making behaviour). Proc Biol Sci 278, 1203–1208. https://doi.org/10.1098/rspb.2010.1607

Zerbolio, D. J., & Royalty, J. L., 1983. Matching and oddity conditional discrimination in the goldfish as avoidance responses: Evidence for conceptual avoidance learning. In Animal Learning & Behavior (Vol. 11, Issue 3, pp. 341–348). https://doi.org/10.3758/bf03199786