REVIEW

Decision-making under uncertainty: biases and Bayesians

Pete C. Trimmer · Alasdair I. Houston ·
James A. R. Marshall · Mike T. Mendl ·
Elizabeth S. Paul · John M. McNamara

Abstract Animals (including humans) often face circumstances in which the best choice of action is not certain. Environmental cues may be ambiguous, and choices may be risky. This paper reviews the theoretical side of decision-making under uncertainty, particularly with regard to unknown risk (ambiguity). We use simple models to show that, irrespective of pay-offs, whether it is optimal to bias probability estimates depends upon how those estimates have been generated. In particular, if estimates have been calculated in a Bayesian framework with a sensible prior, it is best to use unbiased estimates. We review the extent of evidence for and against viewing animals (including humans) as Bayesian decision-makers. We pay particular attention to the Ellsberg Paradox, a classic result from experimental economics, in which human subjects appear to deviate from optimal decision-making by demonstrating an apparent aversion to ambiguity in a choice between two options with equal expected rewards. The paradox initially seems to be an example where decision-making estimates are biased relative to the Bayesian optimum. We discuss the extent to which the Bayesian paradigm might be applied to the evolution of decision-makers and how the Ellsberg Paradox may, with a deeper understanding, be resolved.

It is not certain that everything is uncertain.
Blaise Pascal

P. C. Trimmer
Department of Computer Science, University of Bristol,
Merchant Venturers Bldg, Woodland Rd,
Bristol BS8 1TW, UK

P. C. Trimmer (✉) · A. I. Houston
School of Biological Sciences, University of Bristol,
Bristol, UK
e-mail: trimmer@compsci.bristol.ac.uk

J. A. R. Marshall
Department of Computer Science / Kroto Research Institute,
University of Sheffield, Sheffield, UK

M. T. Mendl · E. S. Paul
School of Veterinary Science, University of Bristol,
Bristol, UK

J. M. McNamara
School of Mathematics, University of Bristol,
Bristol, UK

Introduction

Understanding how animals, including humans, deal with uncertainty is a central issue in the study of decision-making. Experimental studies of decision-making under uncertainty are the focus of a variety of different areas of cognition research including risk-sensitive foraging (e.g. Kacelnik and Bateson 1996; Hayden and Platt 2009; Kawamori and Matsushima 2010), metacognition (e.g. Hampton et al. 2004; Basile et al. 2009; Smith et al. 2010; Call 2010), and neuroeconomics (e.g. Platt and Huettel 2008; Christopoulos et al. 2009). There is also growing empirical interest in how animals respond to ambiguous stimuli that predict either positive or negative outcomes, and whether such responses are 'biased' by the individual's affective state (e.g. Harding et al. 2004; Burman et al. 2008; Mendl et al. 2009). Several authors (Hertwig et al. 2004; Weber et al. 2004; Hayden and Platt 2009; Lagorio and Hackenberg 2010) report that humans show similar behaviour to other animals when faced with similar decision-making tasks.

Alongside these empirical studies, there is a need for theoretical analyses of decision-making that clearly define what is meant by 'uncertainty', identify and formalise how animals might estimate the probability of successful choices under uncertainty and demonstrate whether such estimates underpin optimal decision-making. The aim of this paper is to provide a theoretical basis for thinking about how animals have evolved to make choices under uncertainty, and thus address whether animals should bias their probability estimates when making decisions (as suggested by Nettle 2004).

We assume that decision-making processes have been shaped by natural selection; this allows particular tasks to be analysed using simple models. For instance, optimal foraging choices may be based on expected energy and expected time (e.g. Stephens and Krebs 1986; McNamara and Houston 1997), with additional complexity being introduced by incorporating the risk of predation.

In some situations, it can be worthwhile for an animal to collect more information to reduce uncertainty. Houston et al. (1980) and Erichsen et al. (1980) show that the behaviour of great tits (*Parus major* L.) can be explained by including the time required to identify different types of prey in a model based on the rate of energetic gain. Langen (1999) shows that scrub-jays (*Alphelocoma californica*) are able to modify their food assessment strategies based upon repeated interactions with peanuts of differing qualities. McNamara and Houston (1985) show how information gained during foraging (such as the time taken to crack a nut) can be used to maximise expected reward rate.

To fully incorporate uncertainty into models of animal decision-making, we first need to clearly define what is meant by the term. Knight (1921) distinguished two aspects of uncertainty: determinate and indeterminate (or unmeasurable) uncertainties. These aspects have come to be termed risk and ambiguity, respectively (Ellsberg 1961).

With pure *risk*, for a given choice, the probability of each possible outcome is known. In such circumstances, the best course of action is relatively easy to calculate when the expected value of each outcome is also known. In behavioural ecology, each individual is assumed to maximise its expected reproductive success. An extensive theory of risk-sensitive foraging has been developed to study the expected value of options with differing risks and energetic gains. For examples, see Houston and McNamara (1982, 1985), McNamara and Houston (1987), McNamara et al. (1991), Bednekoff and Houston (1994); for reviews see Real and Caraco (1986), McNamara and Houston (1992) and Houston and McNamara (1999).

With *ambiguity*, the probability of each outcome is not known. One way to deal with such scenarios is to view each situation as consisting of a distribution of probabilities; each probability relating to a particular hypothesis regarding the true probability (Camerer and Weber 1992). The subject of decision-making under ambiguity has received considerable interest in neurophysiology (for instance, see Platt and Huettel 2008; Schultz et al. 2008; Bach et al. 2009), and our understanding of the relevant mental processes may have important clinical implications because the way in which ambiguous information is interpreted can both reflect and generate affective disorders such as depression and anxiety (e.g. Blanchette and Richards 2003; Mogg et al. 2006).

Approach and aims

In this paper, we consider decision-making involving choosing between various options, each of which could result in a successful or unsuccessful outcome. We assume that the value of each outcome is known and focus on decision-making in ambiguous circumstances. Without knowledge of the probability of success of each option, the value of each outcome is insufficient to optimise a decision-making process. In such circumstances, the decision-making problem of choosing between alternatives is ill-defined, as there is not enough information to solve the problem.

The performance of a decision procedure can only be assessed if a distribution of the possible probabilities of success exists. If this underlying distribution of possible values is known by a decision-maker, the Bayesian approach of using this as an initial (or 'prior') distribution of possible values and updating it with new information can be used to make optimal decisions (e.g. McNamara and Houston 1980). If the underlying distribution is not known, it may be possible to approximate it.

In a biological scenario, natural selection will act upon decision-makers according to the actual distribution of outcome values. In Section 'The Bayesian perspective', we describe how Bayesian estimation occurs through a process of updating beliefs. Any organism that is able to approximate both the initial distribution of possible values and the Bayesian updating procedure can be expected to outperform others with less accurate approximations to the optimal strategy. Therefore, organisms behaving *as if* they use Bayesian procedures will tend to be favoured by natural selection. The extent to which behaviour in a given circumstance corresponds to the Bayesian optimum should depend upon the selection pressures imposed by such circumstances in an organism's phylogenetic history and the updates during its lifetime.

To clarify this position, we identify how probabilities should be estimated under particular circumstances and whether biases should be applied to different types of estimation before making a decision in ambiguous circumstances. We then discuss the extent to which the

Bayesian perspective can be applied to problems involving ambiguity, such as the Ellsberg Paradox (which we summarise in Section 'The Ellsberg Paradox').

A simple scenario

To consider the effects of estimations and biases in decision-making, we turn to a simple scenario discussed by Nettle (2004), in which an animal has the choice of whether to attempt something or not. If the animal chooses not to act, it receives zero pay-off. If it acts and succeeds, it gains a benefit b. If it acts and fails, it pays a cost c.

It is assumed that the pay-offs are known to the animal and that only one choice will be made, with no future chance of making such attempts. Thus, there is no value in gaining information for future decisions.

If the actual probability of success, p_a, is known, the expected gain is maximised by choosing to act only if $p_a b \geq (1 - p_a)c$. Thus, the critical value of the probability of success is $p_r = \frac{c}{b+c}$. For $p_a > p_r$, the animal should act, otherwise not.

By assuming that the value of p_a is not known to the decision-maker but can be estimated from previous experience, we consider how p_a should be estimated and whether it is better to bias the estimate, p_{est}, towards or against acting when comparing p_{est} with p_r.

Nettle suggests that it is better to bias towards acting if the benefit of success outweighs the cost of failure, because there is a greater range of p_a for which it is better to have acted than not when $b > c$: 'if information is imperfect, an overestimate would be better than an unbiased estimate of p'. We consider whether this statement is true in the next three sections by considering different approaches to estimation and their resulting effect upon optimal bias.

Frequentist estimation

Assume that p_a is estimated after having witnessed k successes from n trials. The (unbiased) frequentist estimate of the probability of success is k/n. This is also known as the *maximum likelihood estimate*, because the probability (or in hindsight, likelihood) of obtaining k successes from n trials is highest for $p_a = k/n$.

If the estimated probability of success, p_{est}, is used as if it is the true probability of success, then in the case of an unbiased frequentist estimate, the decision to act will occur if $\frac{k}{n} > \frac{c}{b+c}$.

By introducing a bias to the frequentist estimate and regarding p_a as having been chosen from a uniform distribution over the interval (0,1), we can determine whether

it is best to use an unbiased estimate. In 'Appendix 1', we calculate the optimal bias for a range of n and b/c. Using a frequentist (maximum likelihood) estimate, it is preferable to bias optimistically when $b > c$. As the amount of experience increases and the estimate improves, the optimal bias reduces towards zero.

This result implicitly assumes that any value of p_a in the region [0,1] is as likely as any other. The situation can therefore not be regarded as completely ambiguous, because the value of p_a is as equally likely to lie between, say, 0.4 and 0.5 as it is along any other interval of the same length, such as 0.9 and 1.0. The optimal bias to the estimate of k/n will depend upon how the possible values are distributed.

Fixed Errors

We now consider a method of estimation which we regard as unrealistic to highlight the point that without background information, problems may often be ill-defined.

Imagine that an animal is supplied an estimate of p_a with a fixed error, either as $p_a + \delta$ or $p_a - \delta$, but does not know the value of δ, nor has any information from which to estimate its value. Given that the animal will compare p_{est} with p_r to decide whether to act or not, would it be better to have been supplied (i.e. would it be better to *use*) an estimate of $p_a + \delta$ (biasing towards acting) or $p_a - \delta$ (biasing against acting)?

Bouskila and Blumstein (1992) analyse an almost equivalent scenario, in which a foraging individual has to avoid death from starvation and predation. The forager may overestimate or underestimate its risk of predation by a fixed percentage. Bouskila and Blumstein conclude that individuals should err on the side of overestimating the risk of predation. This is based upon a model using rules of thumb for risk assessment (i.e. using imperfect information), and looking at the 'zone of no effect', within which performance is equivalent to having perfect information (i.e. an accurate knowledge of risk). As the zone was typically shifted towards the overestimation of risk, Bouskila and Blumstein claimed that it is generally better to overestimate the risk of predation when circumstances are not accurately known; i.e. individuals should appear pessimistic in terms of predation risk. Their scenario assumed that $b < c$, as the cost of predation is greater than the benefit of food gain without predation. Thus, Bouskila and Blumstein's result agrees with Nettle's claim that individuals should tend to be biased towards taking action.

However, Abrams (1994) uses an equation which allows for different starvation-versus-predation relationships to show that Bouskila and Blumstein's claim is not true in general; whether it is better to underestimate or overestimate

predation risk depends upon particulars of the circumstance faced by the forager.

An approach based on fixed errors may have little or no biological significance; it is difficult to conceive of a biologically realistic scenario in which estimates would take such a form. Rather than regarding errors as fixed offsets from the underlying truth, it is important to represent the method of estimation and thereby induce the consequent errors. The method of estimation can affect whether an estimate is optimal or could be further improved by incorporating a bias in a particular direction.

More fundamentally, without any knowledge of the probability that the supplied estimate is biased positively or negatively, the problem is not well defined. As stated by McNamara and Houston (1980),

> Given a family of probability models […], one must have some prior knowledge about the likelihood that any particular [model] represents the true state of nature. Without such knowledge it is not possible to talk about maximizing expected rewards and therefore impossible to talk about optimality.

The incorporation of prior expectations about the situation allows such problems to become well defined. This mindset accords with the Bayesian perspective.

The Bayesian perspective

In this section, we provide a brief introduction to Bayesoptimal decision-making; a glossary of terms is provided in Table 1. Let us assume that we are interested in estimating the probability of a particular event, A. Without any information about the occurrence of other events, our best guess about the probability of event A is $P(A)$, which is known as the *prior* probability. If another event, B, occurs, then we are interested in the *conditional probability* of event A given that B has occurred, $P(A|B)$. This is known as the *posterior* probability.

If we know the conditional probability of event B given that A has occurred (known as the *likelihood*), then Bayes' theorem allows us to calculate the posterior probability according to:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (6.1)$$

Bayes' theorem therefore allows us to update our knowledge as events occur, from a prior probability $P(A)$ to a posterior $P(A|B)$, where the value of the posterior distribution at any point is proportional to the product of the prior and the likelihood (the denominator normalises the right-hand side of Eq. 6.1).

We now return to the decision problem introduced in Section 'Frequentist estimation'. To accord with the Bayesian decision-making framework, we assume that each individual has a prior distribution of possibilities (in this case, a distribution of the possible values of p_a) and that each update may modify this underlying perception.

The calculation of optimal bias for the frequentist estimate relied upon the assumption that each value of p_a was equally (i.e. uniformly) likely, denoted $p_a \sim U(0,1)$. Taking a Bayesian perspective on the estimation of p_a in such a scenario, the prior would reflect the uniform likelihood of each possible value of p_a. A uniform prior distribution corresponds to the probability distribution of p_a and is typically referred to as an *uninformative prior* (see Table 1), but it does not represent a complete lack of knowledge, as each interval of equal length is assigned the same probability of containing p_a. This prior distribution could represent the true distribution, or an assumption (or estimate) on the part of the decision-maker.

Repeated application of Eq. 6.1 can be difficult; the problem is simplified when the prior distribution is from a family for which application of Bayesian updating also results in a distribution from the same family, which requires minimal calculation. Such families are known as *conjugate priors* (Raiffa and Schlaifer 1961; see Table 1), one example of which is the family of *beta distributions*. These are summarised in 'Appendix 2', where we show that a uniform prior can be represented using a specific beta distribution and that updates can be made repeatedly by treating the old posterior estimate as the new prior before each update.

The mean of the posterior distribution provides the expected value of p_a as $E(p_a) = \frac{k+\alpha}{n+\alpha+\beta}$, where α and β are hyperparameters (see Table 1) of the prior beta distribution. The beta distribution is uniform when $\alpha = \beta = 1$. For a uniform prior, the mean of the posterior after k successes from n trials therefore supplies the expected value of p_a as $(k + 1)/(n + 2)$. This is sometimes referred to as the 'Laplace correction' of the frequentist (maximum likelihood) estimate and converges on the unbiased frequentist estimate for large n.

In 'Appendix 3', we show that it is best not to apply any bias to this expected value before comparing with p_r; i.e. $p_{est} = E(p_a)$ is the best estimate to use. If the prior used in the calculations corresponds to the actual distribution from which p_a was drawn, then no matter what distribution is used to study bias in the context described in Section 'A simple scenario', the Bayesian approach will be optimal with zero bias.

We are now able to see when the frequentist estimate tends to under- or over-estimate the true probability of success. Assuming a prior distribution in the form of a beta

Table 1 A glossary of technical terms used in frequentist and Bayesian estimation

Terminology	Description
Bayes estimate	An estimate of some parameter, such as the bias of a coin, calculated using *Bayes' theorem*. In applying Bayes' theorem to estimation and decision-making, the posterior distribution is thus a *conditional probability* distribution over the possible values of the parameter of interest
Bayes' theorem	A classic result from probability theory, showing how a *posterior probability distribution* can be calculated from evidence available (using a *likelihood function*) and previous information on the probability of different outcomes (the *prior* distribution). See Eq. 6.1
Conditional probability	Denoted $P(A|B)$, this is the *probability distribution* of A, given available evidence B. Conditional probability can be calculated using Bayes' theorem
Conjugate prior	A *prior* distribution belonging to a family of distributions which, when used as the prior for Bayes' theorem, result in *posterior* distributions that are members of the same family. Calculation of Bayes' theorem with conjugate priors is mathematically much simpler than the general case, amounting to update of the distribution's *hyperparameters*. Conjugate priors also allow repeated estimation with Bayes' theorem by using the calculated posterior as a new prior, and combining this with new evidence
Frequentist estimate	A non-Bayesian estimate, calculated solely on the basis of observed event frequencies without taking account of *prior* information. Also referred to as the *maximum likelihood estimate*
Hyperparameter	A parameter that controls the shape of a *prior* distribution
Likelihood	The retrospective probability that an observed event could have happened given a particular parameter value. For example, the likelihood that a coin is fair given that nine out of ten tosses came up heads
Likelihood function	A function that calculates the *likelihood* that a particular value of a parameter in a stochastic model could have generated the observed evidence
Maximum likelihood estimate	The estimate of a parameter as the value having the highest *likelihood*. See also *frequentist estimate*
Probability distribution	For some parameter, a distribution specifying how probable each possible value of that parameter is. See also *prior*, and *posterior*
Prior	A *probability distribution* over different possible states of nature, possibly assumed, or updated using Bayes' theorem given observed evidence and an initial assumed prior
Posterior	The *probability distribution* over possible outcomes that result from the application of Bayes' theorem to a *prior* and some available evidence. In estimation and decision-making, a single value of the parameter of interest is typically calculated as the mean of the *posterior* distribution
Uninformative prior	A prior chosen not to prefer any particular value, such as a uniform prior

distribution, the frequentist estimate is less than the Bayesian estimate when:

$$\frac{k}{n} < \frac{\alpha + k}{\alpha + \beta + n},$$

which reduces to

$$\frac{k}{n} < \frac{\alpha}{\alpha + \beta},$$

the right-hand side of which is the mean of the prior distribution.

Thus, for a beta-distributed prior, the mean of the posterior distribution (the Bayesian estimate) will always be offset from the frequentist estimate towards the mean of the prior. If the variance of the prior is small (i.e. α and β are large, see Fig. 1b), the posterior estimate will tend to coincide with the prior. As the variance of the prior increases (α and β get close to zero, see Fig. 1a), the mean of the posterior tends towards the maximum likelihood estimate (i.e. the frequentist estimate).

With a uniform prior, the mean of the prior is 0.5, so the frequentist estimate is always expected to over- (under-) estimate when the number of successes has been greater (less) than the number of failures.

To see the effect of the prior distribution, let us assume that in a task involving only success or failure, 4 trials have been witnessed of which 3 were successful. If someone with little idea of the probability of success (except knowing that the probability was not 0 or 1) were to use a relatively uninformative (i.e. evenly spread) prior distribution, such as the prior shown in Fig. 1a, then witnessed trials would have a significant effect upon the posterior distribution, as shown. However, if the task were the tossing of a coin which looks fair, with success being denoted by a head, there would be good reason to use a prior with little variance around 0.5, as shown in Fig. 1b. Updating this distribution to the posterior distribution by the use of the same likelihood function (see Table 1) would then have relatively little effect upon the estimate of success. Thus, although the expected value of each prior is the same, the distribution of prior values has a marked effect upon the expected probability of success after witnessing the same sequence of trial outcomes.

Fig. 1 The effect of prior variance on posterior distribution. The likelihoods have been normalised to assist readability; they are the same in each case. **a** Uses a prior distribution with beta parameters $\alpha = \beta = 1.5$. **b** Uses a stronger prior (i.e. with a smaller variance) with beta parameters $\alpha = \beta = 20$
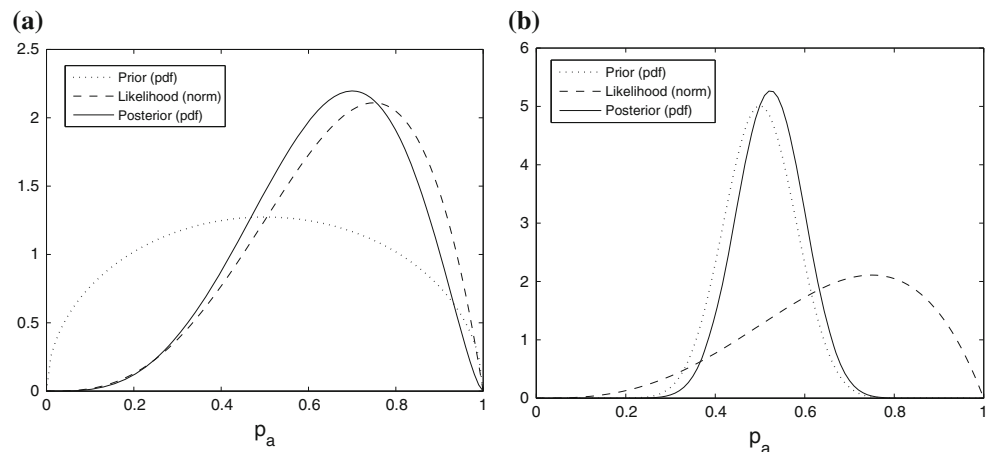
Comparing the results of the fixed error, frequentist and Bayesian methods of estimation, we conclude that the claim (Nettle 2004), that it is better to bias towards acting in the face of uncertainty when $b > c$, is ill-defined because the method of estimating probabilities (i.e. frequentist, Bayesian, etc.) is not specified. Instead, the posterior estimate (or in the absence of data, the mean of the prior) should be compared directly against the critical probability, p_r, without bias.

This conclusion is based upon the assumption that no additional information will be gained (by the choice of either option) for related decisions in the future. Sometimes, one might expect that in such a situation, choosing not to play would result in no additional information whereas playing would improve the accuracy of the estimate of p_a for future decisions. In this case, a 'bias' towards playing should be introduced—though the full Bayesian treatment of sequential decision-making includes possible future outcomes, so any such bias would disappear in the full Bayesian treatment.

We have identified that Bayesian estimates (with sensible priors) should be used without bias; this focuses attention on the question of how humans and animals generate estimates in daily life. For a general account of Bayesian priors and updates in relation to animal behaviour, see McNamara and Houston (1980). For further discussion and examples, see Olsson et al. (1999), Klaassen et al. (2006), McNamara et al. (2006) and Dayan and Daw (2008).

A study may indicate that an animal's decision-making process relies upon periodic updates; based on the above, it may be natural to assume that those updates are of an approximately Bayesian nature (e.g. Naug and Arathi 2007). Valone (2006) reviews empirical data for whether various animal species behave as though they use Bayesian updates, concluding that in 10 of the 11 reviewed cases

they appear to do so. If our behaviour approximates the Bayesian optimum, then evolution may have given us reasonable priors for situations which were faced regularly in our ancestral lifestyles.

When a situation lies outside a subject's evolutionary background, it may be very difficult to predict the relevant beliefs (i.e. priors). Further, an argument that we are 'Bayesian machines' would be overly simplistic. Vast computations would be required to constantly update distributions of probabilities in this world of uncertainty and constant sensory data (see Oaksford and Chater 2009). Lange and Dukas (2009) show that approximations to the full Bayesian treatment can produce similar levels of performance but with only a fraction of the computational requirements. It is more likely that we use rules of thumb (see McNamara and Houston 1980; Stephens and Krebs 1986; Gigerenzer and Todd 1999; Houston 2009) and the optimal bias will therefore depend upon how the rules of thumb generalise. When such rules of thumb are used in unusual circumstances, the resulting behaviour may not correspond to the optimal (i.e. Bayesian) estimates and biases may become apparent through behaviour; we discuss this possibility further in the next section.

Since optimal decisions should depend upon priors, the relevant question in each situation is, 'What is the view of the subject?' It is well recognised that a subject may often have a different perspective on a laboratory situation to an experimenter (see, for instance, McNamara and Houston 1980; Houston and McNamara 1989). Thus, both the power and potential difficulties of applying the Bayesian approach tend to depend upon the use of suitable prior distributions.

To illustrate the flexibility of the Bayesian approach and to highlight some of the potential confounding factors and the care which should be taken when applying such techniques, we turn to a paradox about which it has been said that the Bayesian approach fails.

The Ellsberg paradox

Ellsberg ([1961](#)) interprets the results of various experiments as indicating an aversion to ambiguity. For instance, in an experiment designed to contrast risk aversion with ambiguity aversion, individuals are presented with two urns, each containing 100 balls. Urn 1 contains 50 red and 50 black balls. As the probability of selecting a ball of a particular colour at random is known, urn 1 is termed the 'risky' option. Urn 2 contains between 0 and 100 red balls with the remainder being black; this is the 'ambiguous' option. If asked to choose an urn, with a known monetary reward if a randomly selected ball from that urn is red, the majority of individuals choose urn 1, the risky urn. Equivalent results hold with colours exchanged. Other similar tests, such as a three-colour version of the test (Ellsberg [1961](#)), also indicate that individuals tend to be more ambiguity averse than risk averse.

A simple analysis of the situation shows that each urn has the same expected value, i.e. there is no reason to favour either urn. As people tend to show a preference for the risky urn, Ellsberg concludes that the Bayesian approach 'gives wrong predictions, and by their lights, bad advice.' Other studies (e.g. Viscusi and Magat [1992](#)) conclude similarly: 'There is strong evidence of ambiguous belief aversion, even after one takes into account the full ramifications of a Bayesian learning process', and, 'There is an additional component to the choice process involving ambiguous risks that cannot be reconciled within a Bayesian decision framework.' However, we assert that the prior beliefs being used by individuals may govern the choice of urn within a Bayesian framework.

Ellsberg comes remarkably close to our views when identifying several possibilities for the sort of thoughts which may induce an individual to prefer the risky option to the ambiguous option. For instance, Ellsberg ([1961](#), p. 658) writes:

> Even in our examples, it would be misleading to place much emphasis on the notion that a subject has *no* information about the contents of an urn on which no observations have been made. The subject can always ask himself: "What is the likelihood that the experimenter has rigged this urn? Assuming that he has, what proportion of red balls did he probably set? If he is trying to trick me, how is he going about it? What other bets is he going to offer me? What sort of results is he after?" If he has had a lot of experience with psychological tests before, he may be able to bring to bear a good deal of information and intuition that seems relevant to the problem of weighting the different hypotheses.

It is these very intuitions and biases in background beliefs which we believe would need to be incorporated into the prior probabilities of distributions in order to make an accurate estimation of behaviour. When taking the naive view that no other background beliefs are used and that each possibility for the number of red balls in the ambiguous urn is equally likely, each option is equal in expected value. Consequently, only a tiny discrepancy need be introduced by other assumptions for a bias in choice to appear in an optimal decision-making framework.

The simplest and most fundamental way in which Ellsberg's result may emerge is through individuals not regarding each possibility (e.g. for number of red balls in the ambiguous urn) as equally likely, perhaps through distrust of the experimenter. In some game-theoretic scenarios, such as the Prisoner's Dilemma and the Ultimatum game, the behaviour of subjects has been found to differ depending upon whether they are interacting with another human or a machine (Rilling et al. [2004](#)). The possibility of a game-theoretic explanation is discussed at greater length by Al-Najjar and Weinstein ([2009](#)) who note that, 'we almost always find ourselves playing against opponents with the ability to change the odds.' In the Ellsberg Paradox, if the subject's prior beliefs about the mindset of the experimenter are taken into account (e.g. an expectation of malevolence; see Ozdenoren and Peck [2008](#)), the existence of the choice offered, in itself, may affect the prior expectations of urn values. Al-Najjar and Weinstein conclude that although the Ellsberg Paradox could be explained by an aversion to ambiguity or by the misapplication of heuristics, an assumption of ambiguity aversion can only be sensible in a game-theoretic scenario. An analysis of further options, such as misunderstanding the task slightly, rewards being modified by the perceptions of others (i.e. fear of negative evaluation by others) and decisions being made for several choices at once, can be found in Roberts ([1963](#)), Dobbs ([1991](#)), Trautmann et al. ([2008](#)) and Halevy and Feltkamp ([2005](#)).

Pulford ([2009](#)) found empirical evidence for optimists (as measured on the Extended Life Orientation Test, ELOT) showing significantly less ambiguity aversion than others. The study also showed a significant effect of trust, as the difference between optimists and others was less significant when the contents of the ambiguous urn could have been subject to tampering.

By training rhesus monkeys (*Macaca mulatta*) to respond to visual stimuli for juice rewards, Hayden et al. ([2010](#)) show that the ambiguity aversion of the Ellsberg Paradox applies also to non-human animals. By continuing to subject two of the subjects to ambiguous stimuli, Hayden et al. also show that the degree of ambiguity aversion (as opposed to risk aversion) gradually decreases towards zero, as the monkeys gradually learn the underlying probabilities of ambiguous options. Ambiguity aversion has also been shown to be stronger than risk aversion in chimps (*Pan*

troglodytes) and bonobos (*Pan paniscus*), in tests of food preference (Rosati and Hare 2011).

Finally, the extent to which individuals are generally ambiguity averse can be questioned. For instance, Rode et al. (1999) show that individuals take account of both expected outcome and outcome variability when choosing between ambiguous and risky options. They conclude that these factors, in conjunction with the individual's needs, govern the outcome and that individuals 'do not generally avoid ambiguous options'.

Discussion

We have considered whether, in the face of imperfect information, it is better to overestimate the probability of success (i.e. bias behaviour towards taking action in the face of uncertainty) when the benefit of success is known to outweigh the cost of failure. We have shown that the problem, as stated, is ill-defined because the method of estimating probabilities is not specified. If a Bayesian estimate (the mean of the posterior or, in the absence of data, the mean of the prior) is used with a sensible prior, it should be compared directly against the critical probability of success without bias.

Due to the optimal nature of Bayesian decisions, it is often sensible to suppose that the process of natural selection will have resulted in organisms whose behaviour appears to approximate that of a Bayesian decision-maker in the situations they often face. For a more thorough discussion of this topic, see McNamara et al. (2006); for a somewhat contrary view and an application of the Hurwicz optimism–pessimism coefficient to the Ellsberg Paradox, see Binmore (2009). Following the approximate Bayesian mindset, apparent fear of the unknown—better the devil you know than the devil you do not—may reflect historical (external) situations or methods of approximating the Bayesian decision-making framework (internal mechanisms).

External situations faced by ancestors (evolutionary history) or previous lifetime experiences (i.e. learning) may affect prior expectations. For instance, faced with a group of caves in which to shelter, an unknown cave may be unknown because each individual who enters it is killed. If an individual suspects this may be the case, there would be a bias towards the options about which more is known (in the Ellsberg case, the risky urn).

In some situations, a bias can be produced by the internal mechanisms which do the approximation (i.e. estimates may not be truly Bayesian). For example, one way for an animal to approximately weigh the choices would be to imagine the scenario repeated several times, to visualise which option would lead to the most satisfactory reward in total. In this slightly modified scenario, Halevy and Feltkamp (2005)

show that the risky urn of the Ellsberg Paradox should be preferred. Another way to approximate the best outcome would be to imagine taking several balls from each urn, again imagining summing their value for a total from each. Although with a uniform prior the expected reward would be the same, the variances would differ, devaluing the ambiguous urn even more than the risky urn if the individual's utility function were risk averse (Hazen 1992).

We have focused on scenarios involving just a single decision in this paper. However, it is often important to consider the effect of current actions on future decisions:

1. Choices may provide information which is useful in the future (see McNamara and Houston 1980). With the possibility of gaining better knowledge of the probability of success, an ambiguous option would appear more attractive than if a single-shot scenario was assumed (see McNamara and Dall 2010). Welton et al. (2003) consider the lifetime success of an individual and show that from the perspective of a single decision, optimal behaviour can appear biased. The so-called bandit problems mathematically capture the need to trade-off information gain against immediate reward when a series of trials will be carried out. For instance, Krebs et al. (1978) show that great tits (*Parus major* L.) are able to approximate the Bayesian optimum over a number of trials when sampling between two foraging patches and van Gils et al. (2003) show that red knots (*Calidris canutus*) are able to trade-off immediate reward for information gain using the so-called potential value assessment rule (Olsson and Holmgren 1998).

2. A subject may expect to face the same decision again on multiple occasions. Houston et al. (2007) show that experiments may violate transitivity from the perspective of a single decision if individuals expect the same choice to be repeated. An experimenter may erroneously assume that the focal individual is making the best choice between two options on a one-shot basis, so if option A is preferred to B ($A > B$) and option B is preferred to C ($B > C$), then A should be preferred to C ($A > C$). However, if the means and variances of pay-offs differ between options and the subject is following a rule that gives the best performance assuming a repeated state-dependent choice between the same pair of options presented, then quite aside from issues of information gain, the choice may appear non-transitive to the experimenter.

Thus, the prior expectation of whether a choice will be repeated can affect the optimal decision, irrespective of whether there is any ambiguity over the pay-offs for different actions in a scenario. If more than one choice may be available in future, then the value of potential information

gain must also be accounted for in the calculation of optimal behaviour.

Bayesian inference can allow us to calculate optimal behaviour; for instance, Uehara et al. (2005) use Bayes' theorem to identify conditions under which mate-choice copying should occur. However, this does not mean that the underlying mechanism of any brain is necessarily 'Bayesian', with different areas devoted to prior probabilities, updates and finding the mean of the posterior (see McNamara and Houston 1980). Kahneman and Tversky (1972) have catalogued numerous experiments from which they conclude, 'In his evaluation of evidence, man is apparently not a conservative Bayesian: he is not Bayesian at all.' However, Gigerenzer and Hoffrage (1995) have shown that the use of probabilities in the phrasing of story problems is less likely to evoke correct Bayesian reasoning than other equivalent formats, such as the use of frequencies or natural sampling methods (see also Zhu and Gigerenzer 2006). It seems reasonable to assume that evolutionary processes have selected for mechanisms which approximately maximise our fitness (Parker and Maynard Smith 1990). In general, we expect simple rules to have evolved which approximate the Bayesian optimum in important circumstances (i.e. circumstances which are faced often or have high fitness consequences; see Houston 2009), but which may sometimes fail, be it through an inappropriate input format (which our ancestors have not regularly faced) or not corresponding to a fitness-significant problem. Jackson et al. (2002) provide empirical evidence for the latter possibility; spiders of the same species (*Portia labiata*) use different, threat-sensitive, strategies depending on their (allopatric) habitat.

Waite (2008) argues that gray jays (*Perisoreus canadensis*) may use heuristics (i.e. rules) that are not rational from a traditional economic perspective. In some circumstances, many rules are able to perform well (e.g. Houston et al. 1982, in relation to bandit problems), so the question of which rules will be selected for will often depend upon both how well those rules generalise and their accuracy in the circumstances in which they are used most often. Therefore, in some circumstances, the rules will be inappropriate and result in suboptimal behaviour. Such side effects can be indicative of the underlying mechanisms; for instance, Waite's (2008) findings prompt him to propose hypotheses for the heuristics used by jays. Combining the view of Bayesian optimality, knowledge of suboptimal behaviour and neurophysiological results may help us to better understand aspects of the functional organisation of our brains (McNamara and Houston 2009).

In this paper, we have not been concerned with the mechanisms that underlie Bayesian decisions. One possibility is that emotions or mood states provide internal state variables which reflect past experiences and hence may act as proxy Bayesian priors that guide decisions (cf. Loewenstein et al. 2001; Loewenstein and Lerner 2003; Paul et al. 2005; Mendl et al. 2010). The value associated with a decision may depend in part upon the time taken to make suitable approximations or comparisons (cf. Trimmer et al. 2008). Raiffa (1961) talks of his subjective experience when tested by Ellsberg, of wanting time to do calculations, even to the point that he would have 'paid a premium' to make consistent choices. In this light, ambiguous scenarios may tend to require more thought to accurately approximate their value and hence be somewhat devalued a priori from the value assigned by an experimenter. In terms of brain mechanisms engaged under ambiguous and risky decision scenarios, some studies indicate enhanced activation of the human orbitofrontal cortex and amygdala, areas involved in the evaluation of reward, when faced with ambiguous as opposed to risky choices (Hsu et al. 2005; Schultz et al. 2008). The emotionally activating properties of ambiguous decisions may, in the paradigm of the Ellsberg Paradox, contribute to the apparent devaluation of the more ambiguous option.

In summary, the evolution of decision-makers relies upon the distribution of possible values (relating to the probability of success) and how that distribution can change with time. This knowledge can be used as a prior by an informed decision-maker. Over time, natural selection should act upon organisms to approximate such background knowledge and, especially in variable situations, the lifetime experiences of each individual may also be used to refine estimates, possibly from very weak priors over many possible hypotheses. Although the logic thus far is straightforward, the intricacies by which such approximation mechanisms operate and have developed over millennia leave a great deal to be discovered.

Appendix 1: optimal bias for a frequentist estimate

If p_a is estimated after having witnessed k successes from n previous trials, the unbiased frequentist estimate of the probability of success is k/n. If the estimated probability of success, p_{est}, is used as if it is the true probability of success, then the expected pay-off for true p_a will be given by

$$\sum_{k:p_{est} > p_r} P(k|n)(p_a b - (1 - p_a)c)$$

where

$$P(k|n) = \frac{n!}{k!(n - k)!} p_a^k (1 - p_a)^{n-k}$$

is the probability of k successes from n trials.

We introduce a bias, s, as the fraction of the distance from k/n upward (downward) towards 1 (0) for positive (negative) values of bias. By regarding p_a as having been chosen from a uniform distribution over the interval (0,1), we can then calculate the minimal optimal bias for a range of n and b/c, as shown in Fig. 2. By minimal optimal bias, we mean the bias closest to zero for which the expected return is maximised; the reason for the qualification is that there is typically a range of values of bias for which the expected return is maximised.

The ridges in the graph are a result of the discrete nature of the number of successes and the number of trials. As the number of trials increases and the resulting estimate improves, the optimal bias reduces towards zero, together with the step changes becoming smaller.

Appendix 2: use of the beta distribution

A beta distribution is a probability distribution on the interval [0,1], the pdf of which is defined by the two shape parameters (*hyperparameters*), $\alpha > 0$ and $\beta > 0$, according to:

$$\text{Beta}(\alpha, \beta) = f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

where $B(\alpha, \beta)$ is the beta function, in this case serving as a normalisation constant.

For $\alpha = \beta = 1$, the beta distribution is no longer a function of x, so provides the uniform distribution.

Due to the Bernoulli nature of the trials, the posterior distribution following an update is also a beta distribution, as are all subsequent updates (Carlin & Louis, 1996). To see that this is so, let $f(k|p) = p^k(1-p)^{n-k}$ denote the likelihood of witnessing k successes from n trials when the probability of success is p. Then, if the prior $\pi(p)$ is a beta distribution, we have:
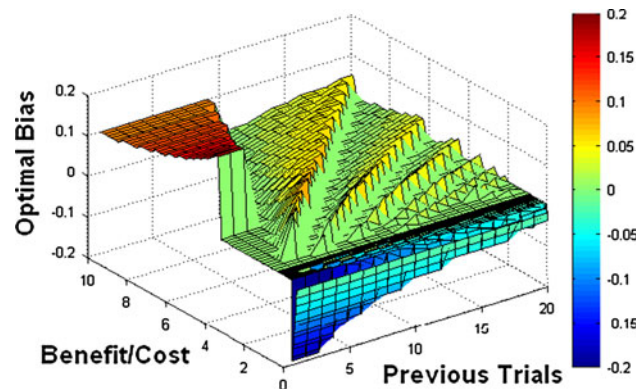
$$P(p|k) \propto f(k|p)\pi(p)$$
$$\propto p^{k+\alpha-1}(1-p)^{n-k+\beta-1}$$
$$\propto \text{Beta}(\alpha', \beta')$$

where $\alpha' = \alpha + k$ and $\beta' = \beta + n - k$.

The mean of the posterior distribution, $\alpha'/(\alpha' + \beta')$, then provides the expected value of p_a as $E(p_a) = \frac{k+\alpha}{n+\alpha+\beta}$.

Appendix 3: comparing the mean of the posterior with the critical probability

The expected probability of success at any stage is given by

$$P(\text{success}) = \int_0^1 f(p)p\,\mathrm{d}p = E(p_a)$$

Using this expected value, it is best not to apply any bias before comparing with p_r.

Let us assume that, having updated the uniform prior with the trial data and obtained a posterior beta distribution with mean p_m, the individual finds that $p_m = p_r$. Will it matter whether the individual chooses to act or not?

If the individual chooses to act, then integrating the expected pay-off across the distribution of possible p values, where $E(p) = p_m = p_r$, the expected value of acting is:

$$\int_0^1 f(p)(pb - (1-p)c)\mathrm{d}p = E(p)b - (1 - E(p))c$$
$$= p_r b - (1 - p_r)c = 0$$

which is the same expected value as not acting. Therefore, if the posterior estimate was greater than the critical probability (i.e. $p_m > p_r$), it would be best to act and vice versa with $p_m < p_r$. The mean of the posterior is therefore a sufficient statistic to compare with p_r when deciding whether to act.

Thus, it is best not to apply any bias before comparing with p_r.

References

Abrams PA (1994) Should prey overestimate the risk of predation? Am Nat 144:317–328

Al-Najjar N, Weinstein J (2009) The ambiguity aversion literature: a critical assessment. Econ Philos 25:249–284

Bach DR, Seymour B, Dolan RJ (2009) Neural activity associated with the passive prediction of ambiguity and risk for aversive events. J Neurosci 29:1648–1656

Basile BM, Hampton RR, Suomi SJ, Murray EA (2009) An assessment of memory awareness in tufted capuchin monkeys (*Cebus apella*). Anim Cogn 12:169–180

Fig. 2 Optimal bias as a function of number of trials and benefit/cost

Bednekoff PA, Houston AI (1994) Dynamic models of mass-dependent predation, risk-sensitive foraging, and premigratory fattening in birds. Ecology 75:1131–1140

Binmore K (2009) Making decisions in large worlds. http://www.else.econ.ucl.ac.uk/papers/uploaded/266.pdf

Blanchette I, Richards A (2003) Anxiety and the interpretation of ambiguous information: beyond the emotion-congruent effect. J Exp Psychol Gen 132:294–309

Bouskila A, Blumstein DT (1992) Rules of thumb for predation hazard assessment - predictions from a dynamic model. Am Nat 139:161–176

Burman OHP, Parker R, Paul ES, Mendl M (2008) A spatial judgement task to determine background emotional state in laboratory rats, *Rattus norvegicus*. Anim Behav 76:801–809

Call J (2010) Do apes know that they could be wrong? Anim Cogn (in press). doi:10.1007/s10071-010-0317-x

Camerer C, Weber M (1992) Recent developments in modeling preferences—uncertainty and ambiguity. J Risk Uncertainty 5:325–370

Carlin BP, Louis TA (1996) Bayes and empirical Bayes methods for data analysis. Monographs on statistics and applied probability 69. Chapman & Hall, London, pp 50–52

Christopoulos GI, Tobler PN, Bossaerts P, Dolan RJ, Schultz W (2009) Neural correlates of value, risk, and risk aversion contributing to decision making under risk. J Neurosci 29:12574–12583

Dayan P, Daw ND (2008) Decision theory, reinforcement learning, and the brain. Cognitive Affective & Behavioral Neuroscience 8(4):429–453

Dobbs IM (1991) A Bayesian approach to decision-making under ambiguity. Economica, New Series 58(232):417–440

Ellsberg D (1961) Risk, ambiguity and the Savage axioms. Q J Econ 75(4):643–669

Erichsen JT, Krebs JR, Houston AI (1980) Optimal foraging and cryptic prey. J Anim Ecol 49:271–276

Gigerenzer G, Hoffrage U (1995) How to improve Bayesian reasoning without instruction: frequency formats. Psychol Rev 102(4):684–704

Gigerenzer G, Todd PM (1999) Simple heuristics that make us smart. Oxford University Press, Oxford

Halevy Y, Feltkamp V (2005) A Bayesian approach to uncertainty aversion. Rev Econ Stud 72(2):449–466

Hampton RR, Zivin A, Murray EA (2004) Rhesus monkeys (*Macaca mulatta*) discriminate between knowing and not knowing and collect information as needed before acting. Anim Cogn 7:239–246

Harding EJ, Paul ES, Mendl M (2004) Animal behavior–cognitive bias and affective state. Nature 427:312

Hayden BY, Platt ML (2009) Gambling for Gatorade: risk-sensitive decision making for fluid rewards in humans. Anim Cogn 12:201–207

Hayden BY, Heilbronner SR, Platt ML (2010) Ambiguity aversion in rhesus macaques. Frontiers in Neuroscience 4:1–7

Hazen GB (1992) Decisions versus policy: an expected utility resolution of the Ellsberg Paradox. In: Geweke J (ed) Decision making under risk and uncertainty: new models and empirical findings. Kluwer, Dordrecht, pp 25–36

Hertwig R, Barron G, Weber EU, Erev I (2004) Decisions from experience and the effect of rare events in risky choice. Psychol Sci 15:534–539

Houston AI (2009) Flying in the face of nature. Behav Process 80:295–305

Houston AI, McNamara JM (1982) A sequential approach to risk-taking. Anim Behav 30:1260–1261

Houston AI, McNamara JM (1985) The choice of two prey types that minimises the probability of starvation. Behav Ecol Sociobiol 17(2):135–141

Houston AI, McNamara JM (1989) The value of food: effects of open and closed economies. Anim Behav 37:546–562

Houston AI, McNamara JM (1999) Models of adaptive behaviour. Cambridge University Press, Cambridge

Houston AI, Krebs JR, Erichsen JT (1980) Optimal prey choice and discrimination time in the great tit *Parus major*. Behav Ecol Sociobiol 6:169–175

Houston AI, Kacelnik A, McNamara JM (1982) Some learning rules for acquiring information. In: McFarland DJ (ed) Functional Ontogeny. Pitman, London, pp 140–191

Houston AI, McNamara JM, Steer MD (2007) Violations of transitivity under fitness maximization. Biol Lett 3:365–367

Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. Science 310(5754):1680–1683

Jackson RR, Pollard SD, Li D, Fijn N (2002) Interpolation variation in the risk-related decisions of *Porta labiata*, an araneophagic jumping spider (Araneae, Salticidae), during predatory sequences with spitting spiders. Anim Cogn 5:215–223

Kacelnik A, Bateson M (1996) Risky theories - the effects of variance on foraging decisions. Am Zool 36:402–434

Kahneman D, Tversky A (1972) Subjective probability: a judgement of representativeness. Cogn Psychol 3:430–454

Kawamori A, Matsushima T (2010) Subjective value of risky foods for individual domestic chicks: a hierarchical Bayesian model. Anim Cogn 13:431–441

Klaassen RHG, Nolet BA, van Gils JA, Bauer S (2006) Optimal movement between patches under incomplete information about the spatial distribution of food items. Theor Popul Biol 70:452–463

Knight FH (1921) Risk. uncertainty and profit. Houghton Mifflin Company, New York, pp 43–46

Krebs JR, Kacelnik A, Taylor P (1978) Test of optimal sampling by foraging great tits. Nature 275:27–31

Lagorio CH, Hackenberg TD (2010) Risky choice in pigeons and humans: a cross-species comparison. J Exp Anal Behav 93:27–44

Lange A, Dukas R (2009) Bayesian approximations and extensions: optimal decisions for small brains and possibly big ones too. J Theor Biol 259(3):503–516

Langen TA (1999) How western scrub-jays (*Alphelocoma californica*) select a nut: effects of the number of options, variation in nut size, and social competition among foragers. Anim Cogn 2:223–233

Loewenstein G, Lerner JS (2003) The role of affect in decision making. In: Davidson RJ, Scherer KR, Goldsmith HH (eds) Handbook of Affective Sciences. Oxford University Press, Oxford, pp 619–642

Loewenstein GF, Weber EU, Hsee CK, Welch N (2001) Risk as feelings. Psychol Bull 127(2):267–286

McNamara JM, Dall SRX (2010) Information is a fitness enhancing resource. Oikos 119:231–236

McNamara JM, Houston AI (1980) The application of statistical decision theory to animal behaviour. J Theor Biol 85:673–690

McNamara JM, Houston AI (1985) A simple model of information use in the exploitation of patchily distributed food. Anim Behav 33:553–560

McNamara JM, Houston AI (1987) A general framework for understanding the effects of variability and interruptions on foraging behaviour. Acta Biotheor 36:3–22

McNamara JM, Houston AI (1992) Risk-sensitive foraging—a review of the theory. Bull Math Biol 54:355–378

McNamara JM, Houston AI (1997) Currencies for foraging based on energetic gain. Am Nat 150:603–617

McNamara JM, Houston AI (2009) Integrating function and mechanism. Trends Ecol Evol 24(12):670–675

McNamara JM, Merad S, Houston AI (1991) A model of risk-sensitive foraging for a reproducing animal. Anim Behav 41:787–792

McNamara JM, Green RF, Olsson O (2006) Bayes' theorem and its applications in animal behaviour. Oikos 112:243–251

Mendl M, Burman OHP, Parker RMA, Paul ES (2009) Cognitive bias as an indicator of animal emotion and welfare: emerging evidence and underlying mechanisms. Appl Anim Behav Sci 118:161–181

Mendl M, Burman OHP, Paul ES (2010) An integrative and functional framework for the study of animal emotion and mood. Proc R Soc B 277:2895–2904

Mogg K, Bradbury KE, Bradley BP (2006) Interpretation of ambiguous information in clinical depression. Behav Res Ther 44:1411–1419

Naug D, Arathi HS (2007) Sampling and decision rules used by honey bees in a foraging arena. Anim Cogn 10:117–124

Nettle D (2004) Adaptive illusions: optimism, control and human rationality. In: Evans D, Cruse P (eds) Emotion, evolution and rationality. Oxford University Press, Oxford, pp 193–208

Oaksford M, Chater N (2009) Precis of Bayesian rationality: the probabilistic approach to human reasoning. Behav Brain Sci 32:69–120

Olsson O, Holmgren NMA (1998) The survival-rate-maximizing policy for Bayesian foragers: wait for good news. Behav Ecol 9:345–353

Olsson O, Wiktander U, Holmgren NMA, Nilsson S (1999) Gaining ecological information about Bayesian foragers through their behaviour. II. A field test with woodpeckers. Oikos 87:264–276

Ozdenoren E, Peck J (2008) Ambiguity aversion, games against nature, and dynamic consistency. Game Econ Behav 62:106–115

Parker GA, Maynard Smith J (1990) Optimality theory in evolutionary biology. Nature 348:27–33

Paul ES, Harding EJ, Mendl M (2005) Measuring emotional processes in animals: the utility of a cognitive approach. Neurosci Biobehav 29:469–491

Platt ML, Huettel SA (2008) Risky business: the neuroeconomics of decision making under uncertainty. Nature Neurosci 11:398–403

Pulford BD (2009) Is luck on my side? Optimism, pessimism, and ambiguity aversion. Q J Exp Psychol 62:1079–1087

Raiffa H (1961) Risk, ambiguity, and the Savage axioms: comment. Q J Econ 75:690–694

Raiffa H, Schlaifer R (1961) Applied statistical decision theory. MIT Press, Cambridge, Massachusetts

Real L, Caraco T (1986) Risk and foraging in stochastic environments. Annu Rev Ecol Syst 17:371–390

Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD (2004) The neural correlates of theory of mind within interpersonal interactions. NeuroImage 22:1694–1703

Roberts HV (1963) Risk, ambiguity, and the Savage axioms: comment. Q J Econ 77(2):327–336

Rode C, Cosmides L, Hell W, Tooby J (1999) When and why do people avoid unknown probabilities in decisions under uncertainty? Testing some predictions from optimal foraging theory. Cognition 72(3):269–304

Rosati AG, Hare B (2011) Chimpanzees and bonobos distinguish between risk and ambiguity. Biol Lett 7(1):15–18

Schultz W, Preuschoff K, Camerer C, Hsu M, Fiorillo CD, Tobler PN, Bossaerts P (2008) Explicit neural signals reflecting reward uncertainty. Phil Trans R Soc B 363:3801–3811

Smith JD, Redford JS, Beran MJ, Washburn DA (2010) Rhesus monkeys (Macaca mulatta) adaptively monitor uncertainty while multi-tasking. Anim Cogn 13:93–101

Stephens DW, Krebs JR (1986) Foraging theory. Princeton University Press, Princeton

Trautmann ST, Vieder FM, Wakker PP (2008) Causes of ambiguity aversion: known versus unknown preferences. J Risk Uncertainty 36:225–243

Trimmer PC, Houston AI, Marshall JAR, Bogacz R, Paul ES, Mendl MT, McNamara JM (2008) Mammalian choices: combining fast-but-inaccurate and slow-but-accurate decision-making systems. Proc R Soc B 275:2353–2361

Uehara T, Yokomizo H, Iwasa Y (2005) Mate-choice copying as Bayesian decision making. Am Nat 165(3):403–410

Valone TJ (2006) Are animals capable of Bayesian updating? An empirical review. Oikos 112:252–259

van Gils JA, Schenk IW, Bos O, Piersma T (2003) Incompletely informed shorebirds that face a digestive constraint maximize net energy gain when exploiting patches. Am Nat 161:777–793

Viscusi WK, Magat WA (1992) Bayesian decisions with ambiguous belief aversion. J Risk Uncertainty 5(4):371–387

Waite TA (2008) Preference for oddity: uniqueness heuristic or hierarchical choice process? Anim Cogn 11:707–713

Weber EU, Shafir S, Blais AR (2004) Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation. Psychol Rev 111:430–445

Welton NJ, McNamara JM, Houston AI (2003) Assessing predation risk: optimal behaviour and rules of thumb. Theor Popul Biol 64:417–430

Zhu L, Gigerenzer G (2006) Children can solve Bayesian problems: the role of representation in mental computation. Cognition 98:287–308