Rational Animals?

Matthew Nudds and Susan Hurley, editors

Oxford University Press

Table of Contents

Part I. Types and Levels of Rationality

Part II. Rational vs. Associative Processes

Part III. Metacognition

Part IV. Social Cognition

Part V. Mind-reading vs. Behavior-reading

Part VI. Behavior in Symbolic Environments

Chapter 2

Meanings of rationality

Alex Kacelnik

Abstract: The concept of rationality differs between psychology, philosophy, economics and biology. For psychologists and philosophers, the emphasis is on the process by which decisions are made: rational beliefs are arrived at by reasoning and contrasted with beliefs arrived at by emotion, faith, authority or arbitrary choice. Economists emphasise consistency of choice, regardless of the process and the goal. Biologists use a concept that links both previous ideas. Following Darwin's theory of natural selection, they expect animals to behave as if they had been designed to surpass the fitness of their conspecifics and use optimality to predict behaviour that might achieve this. I introduce the terms PP-rationality, E-rationality and B-rationality to refer to these three different conceptions, and explore the advantages and weaknesses of each of them. The concepts are first discussed and then illustrated with specific examples of research in bird behaviour, including New Caledonian crows' tool design, hummingbirds' preferences between flowers and starlings' choices between walking and flying. I conclude that no single definition of rationality can serve the purposes of the research community but that agreement on meanings and justifications for each stand is both necessary and possible.

> *"When I use a word," Humpty Dumpty said, in rather scornful tone, "it means just what I choose it to mean-neither more nor less"*
>
> *'The question is," said Alice, "whether you* can *make words mean so many different things."*
>
> *"The question is," said Humpty Dumpty, "which is to be master-that's all."*

2.1. Introduction.

The main questions that concern the contributors to this volume are:

- Are any non-human animals rational?
- What are the character and limits of rationality in animals?
- Are unobservable processes such as reasoning valid causal accounts of behaviour?
- What leads to differences in the kind of rationality exhibited by different species?

These are tough issues in the best of cases, but the real problem, as I see it, is that without a semantic effort we cannot even begin to discuss them: the questions contain words whose meanings cannot be assumed to be shared among those interested in the matter. Even accepting that too much defining inhibits thinking about the real issues, and that (as Humpty Dumpty tells us) definitions are arbitrary, clearly we cannot avoid reflecting on what our central theme, "rationality", means for different authors. Responding to this need, my modest goal here is to discuss some ways in which this polysemous word is and perhaps should be used.

Guided by their differing goals and acceptability criteria, scholars in various disciplines have reached within-field consensus on workable definitions of rationality, and they produce data, reflections, models, theorems, and so on that provide evidence for the presence or absence of rationality and its boundaries as they understand them. These definitions, however, are at best consensual within particular fields. In my experience, a great deal of time is wasted arguing at cross-purposes while holding different understandings of rationality in mind. To mitigate this difficulty, I start by presenting an admittedly idiosyncratic discussion of various conceptions of rationality. In the case of my own field, biology, I will be forced to make a definition up, as none really exists at the moment.

I do not think that it is advisable (or feasible) to use a one-size-fits-all definition. Notions from different fields highlight such different aspects that to propose one overarching definition would be futile because few would follow it. I shall instead subsume all meanings of rationality into three categories, derived from my perception of the main uses in Philosophy and Psychology (PP-rationality), in Economics (E-rationality) and in Evolutionary Biology (B-rationality). I find all these uses necessary and appropriate for specific aims, but as I describe each of them I shall highlight what appears to me to be their virtues and their vices.

2.2. PP Rationality

The *Oxford Companion to Philosophy's* entry for "Rationality" is a good starting point:

> This is a feature of cognitive agents that they exhibit when they adopt beliefs on the basis of appropriate reasons […] Aristotle maintained that rationality is the key that distinguishes human beings from other animals. […] A stone or a tree is non-rational because it is not capable of carrying out rational assessment. A being who is capable of being rational but who regularly violates the principles of rational assessment is irrational. […] Rational beliefs have also been contrasted with beliefs arrived at through emotion, faith, authority or by an arbitrary choice. (Brown 1995, p. 744)

I suspect that this definition would sound acceptable to most non-philosophers, and also, to some extent to contemporary cognitive psychologists (behaviourists may feel more comfortable with what I call 'E-rationality', discussed in the next section). Hence I will use this entry as a working definition of PP rationality. Two features are particularly noteworthy.

First, the emphasis is on *process*, not on outcome. We can separate rational from non-rational beliefs depending on how they were arrived at, rather than according to their contents or the pattern of behaviour that results from them. There is clearly a difficulty in distinguishing 'appropriate' from inappropriate reasons, and the criteria for this distinction are likely to depend on cultural context. For example, to believe that giraffes result from a cross between panthers and camels did count as PP-rational once upon a time because it was based on what were then appropriate reasons. Indeed this belief was held by champions of rationality such as Aristotle and other Greek scholars. This empirically mistaken belief would not qualify as rational today, but no doubt it is rational today to believe in theories that will prove factually wrong as time goes by and science progresses.

Second, PP-rationality is understood not in terms of observable behaviours but of entities such as *thoughts* and *beliefs*. To judge whether behaviour is PP-rational one needs to establish if it is caused by beliefs that have emerged from a reasoning process. To assess the PP-rationality of non-humans, we would have to devise means to expose not just our subjects' beliefs and the processes by which they were arrived at, but also to find a basis for judging whether these processes include 'appropriate' reasons in the sense discussed in the previous paragraph. This makes it very hard to assess whether, for example, a lion is rational, irrational or non-rational. The adherence of cognitive psychologists to this approach is typified by Oaksford and Chater (1998), who point out that, although stomachs may be well adapted to perform their function (digestion), "they have no beliefs, desires or knowledge, and hence the question of their rationality does not arise" (p. 5). This exclusion of stomachs places them in the same rationality bracket as stones or trees, and seems reasonable within this framework, but it raises the question of which definition of rationality is at issue when questions about rationality are raised about the non-human world.

These features would appear to place PP-rationality in a wholly unsuitable position to address our brief. Our focus is on non-human animals, whose thoughts, desires and beliefs are inaccessible in practice and possibly also in principle, and certainly are not the stuff of normal animal research. In dealing with non-verbal subjects, biologists find the notion of using such entities as causes of behaviour problematic, even though the use of some of them are now (after the cognitive revolution) widely accepted. Some kinds of behaviour are best explained by reference to 'concepts' and 'representations' that are observable only indirectly. For instance, if an animal is exposed repeatedly to an interval between two events such as a flash and a food reward, it will later show the same interval between the flash and performing a food-related action. Since the animal produces the interval, it is fair to say that the interval is represented in the animal and causes its behaviour. However, the fact that a representation causes behaviour does not imply that the subject has used reasoning.

The difficulties with PP-rationality are not limited to research with non-human animals. Many processes that give rise to the beliefs held by human subjects are in fact inaccessible to the holders of these beliefs, making it very hard to determine whether a belief has been arrived at on the basis of appropriate reasons. The hundred or so possibilities that chess masters are aware of examining before each actual move are a small subset of the available legal moves (de Groot 1965; Simon and Schaeffer 1992). It is likely that this subset is determined by unconscious processes that delve into the 50000 or so positions chess masters remember, and that choices are often made under the irrational influence of emotional or aesthetic factors without the player being aware of their influence or of their access to her full knowledge base of chess positions. Thus, even if the whole process ends in the belief that a given move is best, and if the player feels that she has arrived at this conclusion by reasoning, the elements that entered into her reasoning process may have been influenced by the kinds of mechanism that the present definition would explicitly exclude from rationality. If, say, the player has acquired a Pavlovian aversion to a given position because she saw it while she had a toothache, then she will play so as to avoid it, and in doing so, she will be influenced irrationally by her knowledge base, though this influence and the active parts of her knowledge base may be unconscious.

I am aware that my concerns apply not just to assessments of rationality but to many other aspects of animal experience including welfare, pain, goal-directed behaviour, theory of mind, and so on, as well as to some aspects of human experience. Nevertheless, I think that within our present focus (rationality in non-humans), PP-rationality is particularly hard to assess. The combined weight of these problems would lead me to exclude PP-rationality from my own research, were it not for my (perhaps PP-irrational) desire to “understand” my avian subjects and my belief that some progress can be made through painstaking experimentation. I shall return to this issue in the last of my empirical examples.

The focus of PP-rationality as I have described it is the rationality of beliefs or of the agents that hold them, or--in the language of cognitive psychologists--the rationality of information processing, rather than the rationality of actions. Yet psychologists, along with economists, are often concerned with the latter. To the extent that action is understood as essentially caused by certain mental processes and beliefs, my preceding comments about PP-rationality apply in similar ways. However, some notions of rationality concern themselves not with the mental processes that lead to beliefs or to behaviour, but with the resulting patterns of behaviour itself, and to this I now turn.

2.3. E-Rationality

Economics is not what it once was. Until some time ago economic theory was a basically consistent set of mathematical models developed from rationality assumptions, but nowadays experimental economics (which studies what economic agents actually do without ignoring deviations from what is expected from them) is booming. This is illustrated by the choice of an experimental economist (Vernon Smith) and a cognitive psychologist (Daniel Kahneman) as Economics Nobel laureates for 2002. But many who concede that the assumption of full rationality is unhelpful in describing the actual behaviour of economic agents nevertheless broadly agree with orthodox economists in what they mean by the word. As part of an insightful discussion of the concept of rationality and the reasons why economists had to introduce it, another Economics Nobel laureate wrote:

> It is noteworthy that the everyday usage of the term “rationality” does not correspond to the economist’s definition as transitivity and completeness, that is, maximisation of something. (Arrow 1986, p. S390)

The ‘something’ to which this definition refers is ‘Expected Utility’. Expected utility maximization is itself characterized mathematically in terms ofthe axioms of completeness and transitivity, and related properties such as independence and regularity (See (Mas-Colell, Whinston et al. 1995) for definitions of these terms). These are all properties related to the internal coherence of the agent's choices. Behaviour that is compatible with expected utility maximization counts as rational; in contrast with PP-rationality, the processes that generate behaviour are not the focus. For present purposes, the essential idea is that expected utility maximization constrains patterns of behaviour to be internally coherent in a certain sense, but places no substantive demands on behaviour, such as that choices lead to wealth accumulation, happiness, biological success or the honouring of commitments.

While this is the prevalent view in economics, there are some important dissenting voices. Critics of the "internal consistency" approach include, besides those mentioned so far, Amartya Sen, the Economics Nobel laureate for 1998. In Sen's view, a person who behaves according to these principles may well be rational, but if his behaviour is unrelated to happiness or other substantive concerns then he must also be a bit of a fool (Sen 1977). He proposes a view of rationality that does take into account the substantive interests of mankind. Although I have sympathy for Sen's perspective on what constitutes a suitable target for policy, I shall use the "internal consistency" definition of rationality under the label of E-rationality for this discussion, both because it is the most widely used definition among economic theorists and because it is readily usable in the context of non-human behavioural research.

In contrast with PP-rationality, E-rationality concerns patterns of action rather than beliefs and the cognitive processes that may cause these actions. E-rationality is about observable behaviour rather than about unobservable or private mental states, and about outcomes rather than processes. What matters are the choices made and how they relate to each other rather than whether they originated from emotional impulses or cold reasoning. Indeed, economically rational agents can be institutions (which do not have private mental experiences) rather than persons, opening the possibility that even plants –and perhaps stomachs- may behave rationally.

Despite the apparent epistemological advantages of focussing on observable outcomes rather than unobservable mental processes, E-rationality faces difficulties of its own. The central concept of expected utility maximization is not by itself accessible to observation, but is defined post-facto as whatever is (consistently) maximised by the observed behaviour of the agent. While defining utility by reference to preferences as revealed in behaviour has logical advantages, it also limits considerably the class of observations that could be accepted as a violation of rationality. As long as observed choices can be interpreted as maximizing some common currency, however exotic or perverse, the choices are rational, under this definition. As Sen (1977) puts it, "you can hardly escape maximizing your own utility". Empirically, very little is ruled out by the mere internal consistency of E-rationality, unless we implicitly revert, as non-economists tend to do, to using an intuitive notion of utility as some property of the agent (viz. desire for wealth) that causes him to express those preferences in his behaviour.

The depth of this empirical difficulty can be easily illustrated by reference to state-dependency. For instance, non-transitive preference cycles (preferring a to b, b to c but c to a) are definite irrationalities, assuming that the subject has remained in a constant state. However, if I choose lamb over ice-cream at 8 pm, ice-cream over coffee at 9 pm, and finally coffee over lamb half an hour later, this is not a serious breach of any principle of consistency because I have changed my state in the intervening period, that is, my preferences are state-dependent and this eliminates the intransitive cycle. The point made by this example is theoretically trivial, but in practice, especially when applied to non-humans, it can lead to misleading conclusions, as I shall explain later.

Another example of the limitations of the revealed preferences approach is provided by lactation. In passing resources to her offspring, a lactating mother incurs a material loss. An observer may think, *prima facie,* that for a rational agent defined as utility maximiser any loss entails disutility and may then wonder if lactation violates utility maximisation, but this would be mistaken. According to the revealed preference approach, the mother's choice to pass resources to her offspring simply means that her utility encompasses the well-being of her offspring. The degree to which the child's and her own consumption combine to define the mother's utility can be directly construed from the allocation of available resources. This has been encapsulated in a 'dynastic utility function' that depends on the utilities and number of children of all descendants of the same family line (Becker and Barro 1988). This approach can bring lactation back into the realm of rational behaviour, but it raises the question of how one determines the utilities of all those involved other than tautologically by their observed choices. Economists sometimes revert to the intuitive notion of using a substantial criterion and identify utility with an objective observable such as consumption, but this is not helpful either, since consumption maximisation for either one agent or a dynasty of them has no external justification as a currency for maximisation.

The contrast between PP-rationality and E-rationality is evident: the former deals with causal processes and with unobservable events such as beliefs while the latter deals with outcomes and with observable actions. An agent can be "rational" in either of these senses while being "irrational" in the other. Biology provides yet another approach, and I discuss this next.

2.4. B-Rationality

Rationality has never been a primary concern for evolutionary biologists, but ideas that relate to it underlie the logic of optimality modelling of decision-making in animals and plants. Furthermore, there is a rich and thriving tradition of contact between biology and economics: evolutionary game theory has made important contributions to economic theory, experimental economists challenge concepts within evolutionary theory, and experimental animal behaviourists test the E-rationality of their subjects (Shafir 1994; Hurly and Oseen 1999; Bateson, Healy et al. 2002; Schuck-Paim and Kacelnik 2002; Shafir, Waite et al. 2002; Fehr 2003). Given this busy exchange, it seems useful to make an attempt at tidying up a definition of rationality from a biologist's viewpoint.

B-rationality is necessarily linked to fitness maximization because it is based on the historical process of evolution of behavioural mechanisms. Evolutionary change is caused by both directional (natural selection) and non-directional processes (genetic drift). Both have some predictability. For instance, random genetic drift results in a predictable rate of accumulation of mutations and this serves as a clock to measure evolutionary distances between species. However, only natural selection generates phenotypic properties that can be anticipated using principles of maximisation of a defined currency. For this reason biological rationality is best examined with natural selection (and hence fitness) at the centre.

Fitness is as central to biological theory as utility is to economic theory, but the two concepts are epistemologically very different. Current notions of fitness are grounded not on the revealed preferences of agents but on the genetic theory of natural selection. Broadly speaking, the fitness of a biological agent is its degree of success (growth as a proportion of the population) relative to that of other agents in the same population. Because fitness is relative and not absolute success, the fitness of an agent is always dependent on the population context and is not an intrinsic property of each agent.

Strictly speaking, agents should be alleles, which are different versions of each gene, but with suitable transformations it is possible to discuss the fitness-maximising behaviour of individuals that carry these alleles, using the "individual-as-maximising-agent" metaphor (Grafen 1999; Grafen 2000). The passage from gene to individual level of analysis is not trivial. As an example, I return to lactation, in which the behaviour of the individual-level agent (the mother) compromises her nutritional state for the sake of another individual (her child). From the gene's point of view, one could argue that each allele in the mother's body that has an influence on her behaviour may also be (with a probability of one half) in the body of the child, and hence the allele as an agent is influencing the passage or resources between an older and a younger version of itself. There will be an optimum level of transfer that has to do with the replicating chances of each of the two versions, and this can be examined and predicted. As mentioned earlier, within the utility framework this can be handled by transforming the mother's utility function into a dynastic version that includes consumption of the resource by self and by her descendants, thus building a function that fits the observed behaviour (Becker and Barro 1988).

The biological alternative is to work out the fundamentals from the point of view of the alleles and the mechanics of population genetics, and then identify the appropriate transformation of the predicted allocation of resources to define a function that the individual ought to maximise. (For a rigorous treatment of the individual-as-maximising-agent see Grafen (1999, 2000).) This concept was developed in detail by William Hamilton (1964) under the name of "inclusive fitness" and it takes into account that alleles sit in more than one body and the dynamics of evolving populations. Because of the complexities of sexual reproduction and incomplete dominance between alleles, the individual as maximising agent analogy is not yet fully tied to the genetic theory of natural selection, but theoreticians are busy working on this.

Meanwhile, it is through this analogy that behavioural researchers from an evolutionary persuasion pursue their program.

In summary: Under B-rationality the behaviour of individuals is a function of the genetic material that guides their development and functioning. Natural selection is a consistent (E-rational?) process that determines the distribution of alleles in populations and hence imposes properties on the genes that predominate in the individuals of each species at any given time. As a consequence of natural selection, alleles shape the behaviour of their carriers in ways that promote their own (the alleles') success. Theoretical population genetics is used to define a function (inclusive fitness) that describes how to see the individual as the maximising agent, and this underlies the logic of optimality in biology. What the individual agent maximises with its behaviour can never be understood without reference to the fact that the same alleles sit in more than one body.

This deceptively straightforward theoretical picture opens two difficult new problems: how the concept of individual-level rationality derived from inclusive fitness relates to those used by philosophers, psychologists and economists and how these ideas can be tackled empirically when studying behaviour of real animals.

As a first approximation to the first issue, I suggest that a B-rational individual can be defined as one whose actions maximise its inclusive fitness. This definition is closer to E- rationality than to PP-rationality in that it emphasizes outcome rather than process, it operates with observable behaviour, and it strongly emphasizes consistency. It differs from the former, however, in that what is maximised by B-rational agents (inclusive fitness) is far more constrained than what is maximised by E-rational agents (utility). As we have seen, inclusive fitness is definable and in principle measurable at genetic level, independently of the subject's choices, while utility is constructed from these choices. The cognitive or emotional processes that may accompany or even cause behaviour are not important for B-rationality: the approach is applied equally to bacteria, oak trees, blue whales and humans in spite of their cognitive differences (but does not apply to stones, because in contrast with living things, they have not been 'designed' by natural selection. In terms of the second problem (testing), the main difficulty with B-rationality as defined above is that it is not explicit about the conditions across which the subject can be an inclusive fitness maximiser. It would be unjustified to expect any living creature to be B-rational under all conceivable circumstances. Having no foresight, natural selection only shapes neural mechanisms on the basis of encountered situations, so that as circumstances change evolved mechanisms fall short of fitness maximization. Since no creature can be expected to be universally B-rational, the concept is only useful when relativised to limited sets of circumstances and limited classes of decisions. Individual behaviour is driven by mechanisms evolved because they induce B-rational behaviour and not by the intentional pursuing of fitness maximisation. These mechanisms may well include submission to emotions, authority, faith and false beliefs.

It is worth mentioning, however, that an alternative view is also tenable. Imagine an organism that computes the consequences of each possible action and then uses reason to act in a way that maximises inclusive fitness. That organism would be globally B-rational. It would also be PP-rational because it arrives at its decisions by reasoning and it would be E-rational because inclusive fitness would be identical to its utility as constructed from revealed preferences. I am not aware of anybody who explicitly defends the existence of such a creature, but in expecting humans to behave under a great variety of present cultural environments according to the maximisation of inclusive fitness, some evolutionary psychologists and behavioural ecologists fall only marginally short of assuming that humans operate in this way.

A number of recent biological publications have used the term “rationality” in an unqualified way that primarily reflects its economic definition, namely with emphasis on self-consistency. Thus, hummingbirds, starlings, jays and honeybees have been charged with irrationality (Shafir 1994; Hurly and Oseen 1999; Waite 2001; Bateson, Healy et al. 2002; Shafir, Waite et al. 2002) because their behaviour violates either transitivity or regularity (a principle of choice that states that the addition of further alternatives to a set of options should never increase the level of preference for a member of the original set (Luce 1977; Simonson and Tversky 1992; Tversky and Simonson 1993)). Counterclaims state that these observations may be compatible with B-rationality when state dependency is considered (Schuck, Pompilio & Kacelnik, under review).

Before I proceed to consider how these three definitions of rationality may apply to birds, it may be helpful to summarize the main points. PP-rationality requires that beliefs or actions be based on reasoning; it focuses on how beliefs or actions are arrived at rather than what they consist of. E-rationality focuses on whether behaviour is consistent in the sense of maximizing some function that is called "utility". Utility maximization is not tied to substantive criteria such as fitness or well-being. It focuses on how an individual behaves rather than on how it has arrived at its preferences. B-rationality is the consistent maximisation of inclusive fitness across a set of relevant circumstances; the under-specification of this set is perhaps the main weakness of B-rationality. It should be clear by now that individuals can certainly be rational in some of these senses while violating the other two. In the next section I go through examples of empirical research in bird behaviour that relate to the topics discussed so far.

2.5. Rational birds? Some experimental tests

B-rationality: Optimal Foraging Theory.

Tests of B-rationality typically start by considering situations assumed to be evolutionarily relevant and then proceed to create models that predict behaviour assuming that the subject chooses the strategy that maximises inclusive fitness among a limited set of options (Kacelnik and Cuthill 1987; Krebs and Kacelnik 1991). Such models may fail to predict actual behaviour in experimental conditions, but this is not surprising because several elements of the model-making process can lead to failure.

In the first place, the situation considered may in fact not be ecologically relevant either at present or in the past, meaning that the choice may not have been encountered often enough during the evolutionary history of the species to have shaped its behavioural patterns.

Next, there is a serious problem with the description of the strategy set that the models draw on. It is unsatisfactory and somewhat circular to claim that an animal performs the best action among those in its strategy set and that if it fails to act adaptively it is because it has not got the best behaviour in its repertoire.Often, however, there is no escape from such claims. This is obvious when we deal with a physiological or anatomical limitation. For instance, behavioural ecologists can use models to predict the optimal choice between walking, jumping and flying for locusts but they can only include the first two options for frogs. Clearly, a frog will not fly because flying is not within its strategy set.

However, and more problematically, this limitation also applies when the constraint is psychological. For instance, starlings will forego foraging gains so as to be close to conspecifics (Vásquez and Kacelnik 2000). Overall, this gregarious drive is adaptive because in nature flocking enhances feeding rate through several mechanisms, including sharing of information about both food location and predation danger (Giraldeau and Caraco 2000; Fernández-Juricic, Siller et al. 2004). When we expose a starling to a situation where the best feeding patch is not in the greatest vicinity to other starlings, we expose a psychological mechanism that stops the starling from foraging in a way that maximises foraging yield. This mechanism may have evolved because on average it yields greater benefits, but it does sometimes stop starlings from satisfying optimality predictions for specific situations.

It is impossible to determine precisely which observed features of a creature should be included in the assumptions of an optimality model and which should be left to be predicted by the model itself. This means that testing B-rationality in practice is more dependent on revealed preferences than biologists like myself would like to admit and that some of the circularities of E-rationality afflict this approach as well.

These aspects of the study of B-rationality are evident in any detailed application of optimality in foraging behaviour, and as an example I present one case in some detail.

B-rationality: To fly or to walk?

Starlings forage sometimes by walking on the ground (poking the soil with their beaks to dig for hidden grubs) and sometimes by taking short flights (hawking small airborne insects). In one study (Bautista, Tinbergen et al. 2001) these two foraging modes were taken as given and optimality modelling was used to examine how the birds chose among them in a laboratory situation. Starlings could work for food by walking or by flying. In each foraging mode a number of trips (walks or flights) between a resting place and a food source were necessary to obtain one reward. In eleven different treatments, the flying requirement was fixed as a number of flights between one and eleven, and the number of walks that made the two foraging modes equally attractive to each bird was found using a titration technique: the number of walks increased or decreased by one depending on whether the animal's previous choice had favoured walking or flying respectively. This number eventually oscillated around a certain value, and this was taken to be the number of walks that were as attractive as the number of flights in that treatment.

The model was based on the fact that hawking yields more frequent captures but is more expensive, because flying uses more energy per unit of time than walking. I describe below the details of the model. Skipping these details should not obscure my main point.

Let S_w and τ_w be respectively the size (in energy units) and involvement time (in time units) per prey from walking, while S_f and τ_f are the size and involvement time from flying. Involvement time includes the times taken to travel to and to consume a prey item. Getting food using each mode involved time resting, travelling and handling the food, but the model uses the average metabolic rates throughout all these components. This average differs among modes. Energetic gain is known to play a major role in this sort of problem, but several metrics are reasonable possibilities. Foraging theorists use either energy gain over time ('Net Rate' and its simplification 'Gross Rate') or energy gains per unit of expenditure ('Energetic Efficiency').

Net Rate for an option *i* is defined as $\gamma_i = \frac{Gain_i - Cost_i}{time_i}$, where gain is the energy content of a capture, cost is the energy spent in procuring it, and time is the involvement time. Gain is the usable caloric content of a prey and cost is the product of the metabolic rate (in energy per unit of time) times the involvement time. Walking and flying yield equal net rates when $\gamma_w = \gamma_f$ as expressed by:

$$\frac{S_w - m_w \tau_w}{\tau_w} = \frac{S_f - m_f \tau_f}{\tau_f} \tag{1}$$

where m_w and m_f are the average metabolic rates in the walking and flying modes respectively. When reward sizes are equal between the two modes (as in the experiment discussed here) the cost in walking time that yields the same rate as a particular flying time cost is given by solving Equation 1 for τ_w:

$$\tau_w = \frac{1}{\frac{1}{\tau_f} + \frac{m_w - m_f}{S}} \tag{2}$$

where S is the common size of both rewards. Gross rate is a simplification in which the negative terms in the numerators of equation 1 (the costs) are ignored, that is treated as if their values were negligible. If this is done and reward sizes are equal, then equation 2 simplifies so that equal gross rates are achieved when $\tau_w = \tau_f$.

Efficiency may be the optimal choice when the main constraint is availability of energy for immediate use, for instance if energy gains do not become immediately usable. When this occurs, it makes sense for an optimal decision maker to maximise the ratio of gains to expenditure regardless of the time involved. With the same notation as before, we can see that Efficiency is equalised between modes when

$$\frac{S_w - m_w \tau_w}{m_w \tau_w} = \frac{S_f - m_f \tau_f}{m_f \tau_f}. \tag{3}$$

Then, with equal reward sizes, we get

$$\tau_w = \frac{m_f}{m_w} \tau_f \ . \tag{4}$$

The predictions of the models and the titration results from the starlings are shown in figure 1. As the figure shows, Net Rate maximisation (Equation 2) is extremely successful in predicting the starlings' choices, while the alternatives (Gross Rate and Efficiency) either underestimate or overestimate preference for walking.

FIGURE 2.1 HERE

This example shows that it is possible to predict animal preferences assuming a consistent (hence E-rational) criterion on the part of the subjects, and that this criterion may be deduced *a priori* from functional considerations, so that B-rationality is tested as well. On the other hand, the example also shows that the *a priori* biological criterion was not inclusive fitness but a proxy, and that this proxy was not unique. At least three alternatives (Gross Rate, Net Rate and Efficiency) were judged possible, and which one drives choice was established by reference to the birds' preferences, thus submitting to some extent to the tautology of utility-based approaches. Nothing in this example links to PP-rationality. There is no reference to the psychological mechanism by which the subjects make choices. There is no suggestion for or against the possibility that the starlings reason their way to choice, and hence the study does not connect with PP-rationality.

E-rationality

The economic concept of rationality, with its emphasis on consistency of choice, offers an interesting and radically different source of inspiration for experiments on animal decision-making. At some level E-rationality may be seen as a corollary of B-rationality, but one that leads to different questions and different experiments. If a subject is a consistent maximiser of inclusive fitness, then it is a consistent maximiser of something, and a consistent maximiser of anything is by definition E-rational. It follows that empirical observations of violations of E-rationality pose problems that biologists need to address. Recent years have seen a proliferation of studies with an emphasis on E-rationality (Shafir 1994; Hurly and Oseen 1999; Waite 2001; Bateson 2002; Shafir, Waite et al. 2002) and here I discuss one such example.

One conceptual violation of E-rationality is found in some forms of preference reversals due to context dependence. Imagine that a subject facing a choice between A and B prefers A in one context and B in another. This would be a sign of E-irrationality that should warrant intensive attention and research by biologists. Precisely such an observation was reported by Hurly and Oseen (1999) in a field study of risk preference by wild rufous hummingbirds *Selasphorus rufus*. These authors offered wild, free-living hummingbirds choices between artificial flowers that differed in the level of variance in amount of sugar solution. There were three kinds of flowers, identified by their colour. Type N (for No Variance) always offered the same amount of nectar, type M (for Medium variance) offered a coefficient of variation in volume of 33.3%, and type H (for High Variance) offered a coefficient of variation of 66.6%. If an animal systematically avoids variance in amount (this is a frequent trend; see Kacelnik and Bateson 1996), it should always prefer the option with lowest variance in any set. The hummingbirds complied with this expectation by significantly preferring N over M, M over H and N over H when facing pairs of alternatives. However, when the birds were offered a choice between the three types of flower presented simultaneously, their maximum preference was for the intermediate level of variance, thus reversing the ranking between medium and no variance. Similar observations have been made in other species and paradigms including honeybees, starlings and jays.

Now, how might a biologist respond to observations of this kind? One option is to accept the violation of E-rationality at face value, including the implied breach of B-rationality, and to conclude that the hummingbirds are indeed poor choosers in this situation. This is perfectly tenable, as it is possible that although the same choice (say lower variance) is always better, in their evolutionary history they may have only very rarely faced simultaneous encounters of more than two kinds of flowers, and hence even if this situation leads to a suboptimal preference, the cost has not been sufficient to select for mechanisms that avoid this malfunction. Another possibility is to reflect on whether an adaptive explanation may exist to account for the observed reversal. In fact, several suggestions exist for why relative preferences between two options may change adaptively when other options are added.

One idea (Houston 1997) is that the decision maker infers (meaning here: behaves as if it infers) what options will be available in the future from the set of options present at the time of choice. If this were so in the hummingbird example, an animal facing two sources, one with no variance and the other with medium variance, would infer that this is what the future has in store, and then it would deploy the best choice for that scenario. On the other hand, an animal facing the same two options within an extended set (after the inclusion of a third alternative with even greater variance) would infer that the future will make this third option available as well, and this may change the nature of the problem sufficiently to alter the relative values between the original alternatives. There is no evidence that this explains the hummingbirds' reversal, but Houston's idea illustrates how apparent violations of B and E rationality may be expressions of a complex context-dependent set of optimal strategies.

A related possibility (Schuck-Paim 2003) looks back into the agent's history rather than forward into how the agent infers future options. This idea postulates that the different contexts during training may result in different energetic states at the time preferences are measured. If this occurs, then the problem faced by the agent has changed, and there is no reason to expect consistency in preference ranking. With a subject in a different state, the benefit accrued from each option may be different. The difference in state of the subject in experiments such as those with hummingbirds may occur because preferences are not exclusive. Since the subject allocates some choices to all available options, then its energetic state reflects the conditions that precede the tests, and animals trained in pairwise choices may be in different energetic states (or have a history of different variance in state) from those trained with simultaneous presentations of three options. If the subject's state is different, then relative preferences may reverse without this being an indication of E-rationality. This remark is similar to the point of the "lamb vs. ice cream vs. coffee" example, in which, because state varied as the meal progressed, reversals did not violate rationality.

In summary, apparent violations of E-rationality are interesting and pose challenges to B-rationality, but it may often be the case that when a B-rational interpretation has been found, the notion of a violation of E-rationality may become unnecessary. In particular, it is crucial to show that when a subject shows inconsistent preferences, these are measured with the subject in the same state. In Houston's interpretation the state of the subject differs because it possesses different information about the future and in Schuck-Paim's idea the state differs because the recent history of the animal has placed it in a different energetic state. If any of these ideas is supported, then the breach of either kind of rationality is explained away.

Once again, in this section I did not refer to the notion of PP-rationality, as whether hummingbirds' actions resulted from a reasoning process or otherwise could not and was not addressed. To tackle the possibility of saying something about this issue, I turn to my final avian example.

PP rationality: New Caledonian crows

In the introductory description of PP-rationality I exposed my pessimism regarding the possibility of testing of PP-rationality in non-humans. This stems from the fact that testing this kind of rationality requires access to the process by which subjects reach beliefs, rather than dealing with some mapping between the material situation of the decision maker and its behaviour. Very worthy attempts have been made to cross this bridge by researchers working on topics such as mind reading, gaze following, and other forms of social inference. In most cases the results raise very difficult problems of interpretation (Povinelli 2000; Visalberghi 2000; Tomasello, Call et al. 2003; Tomasello & Call, this volume; Povinelli & Vonk, this volume). Without any expectation of solving the difficulties, I describe below one example from our work with New Caledonian crows (*Corvus moneduloides*) that goes some way to temper my pessimism.

The case in question (Weir, Chappell et al. 2002) concerns what a crow named Betty did when faced an out-of-reach morsel in a small basket at the bottom of a vertical plastic well. On previous occasions Betty had been provided with two wires, one straight and the other hooked, and had been able to lift the basket with the hooked wire. On a crucial trial, however, she only had available a straight wire. After failing to lift the basket with the straight wire, she took it to a fracture in a nearby plastic tray, wedged the tip there and pulled perpendicularly from the proximal side, bending the wire until it formed a hook. She then returned to the well, retrieved the basket and ate the food. Further observations showed that she could bend wires using several different techniques to achieve functional tools.

Many corvids are capable of being trained to use tools (Powell and Kelly 1977), and New Caledonian crows are consummate tool-makers and users (Hunt 1996; Hunt 2000), but before this observation there had been no report of a bird solving a novel problem with such degree of creativeness. New Caledonian crows make and use hook-like tools from plant material in the wild and in the laboratory (Hunt and Gray 2004; Chappell and Kacelnik in press). However, to our knowledge they normally do not have access to pliable material that can be bent into a shape and preserve it. Betty used her motor skills and knowledge of the principle of hooks to devise a new solution to her problem, and this solution required leaving the site of the problem, finding a suitable crack to hold the wire's tip, and modifying the wire appropriately before returning to finish the task. It is hard to account for all of this without reference to some form of planning ahead, some representation of the problem and its solution and some choice among possible actions leading to the solution. This is what many would call thinking, and with only a little suspension of disbelief might even be accepted as reasoning and as approximating the conditions for PP-rationality.

The attribution of any form of rationality cannot be based on one set of observations, however compelling this set may be. We do not know how domain-general the New Caledonian crows' ability to plan and execute solutions to new problems is. We need to investigate this while bearing in mind that there is evidence that even for humans there is no such a thing as totally domain-independent reasoning abilities (Cosmides and Tooby 1997). In a sufficiently attractive scenario we may hope to show that animals reveal underlying processes that include wide range anticipation and planning. This may help to progress even in the hard field of PP-rationality.

2.6. Conclusions: Are animals rational after all?

My purpose in this chapter is to provide a small degree of clarification in the use of the term (and hence the concept) of rationality, because the term is manifestly interesting and worth researching but by virtue of its being used in various fields has diverse and potentially incompatible interpretations. I constrained my narrative to three notions of rationality, and illustrated them with examples from bird research. I am aware that I did not exhaust the discussion of previous uses of the term nor did I necessarily use the most striking available examples (I realise with some dismay that I did not discuss the achievements of Alex the African Grey parrot (Pepperberg 1999 and this volume). As a minimum palliative to this parochialism, I add a few words about a use of the term rationality that I have not discussed so far.

The notion of "Ecological Rationality" has been fostered through many studies by Gerd Gigerenzer, Peter Todd and their colleagues (Gigerenzer, Todd et al. 1999). This concept refers to human subjects' hypothetical use of thinking rules ("fast and frugal heuristics") that achieve satisfactory solutions under the limited sets of circumstances subjects face in their everyday life. Ecological rationality is thus clearly different from the three definitions of rationality used through this paper.

It differs from PP-rationality in that subjects' are not assumed to arrive at beliefs by logical reasoning, from B-rationality in that it does not focus on the evolutionary or developmental origins of each rule, and from E-rationality in that there is no paramount role for internal consistency among choices. The idea that humans follow simple cognitive heuristics to cope efficiently with the problems they face in real life is close to that of "rules of thumb", which were used in early foraging theory to account for how behaviour that may require complicated calculation to be identified as optimal is performed by insects or birds that could not be expected to perform the calculations. Krebs and McCleery expressed this concept thus:

> It is generally assumed that foraging animals use simple 'rules of thumb' to solve their foraging problems, and that these rules may approximate to the solutions predicted by optimization models. (Krebs and McCleery 1984, p. 118)

This notion was applied to many search problems, including for instance the issue of how birds might find a balance between the advantages and costs of acquiring knowledge (Houston, Kacelnik et al. 1982).

The use of this approach for animals has had mixed success. In fact, there is a dichotomy between work on behaviour that is directly related to fitness and follows genetically pre-programmed rules (as when parasitoid wasps track chemical gradients to localise hosts or female birds use the size of males' song repertoires to choose among suitors) and work on problems where subjects learn a great deal, including the nature of the problem and the parameters of each alternative. For the first kind of problem the rules-of-thumb approach works well, but for problems that are evolutionarily unpredictable and require individual learning, unravelling the mechanisms of preference acquisition seems a better strategy. Learning mechanisms (pre-eminently associative learning) are species-wide and generate a variety of behavioural rules, each tailored to a given agent's individual history. Criteria for rationality are more likely to apply to the learning mechanisms by which animals acquire preferences than to the rules that result from them.

It seems appropriate to end this chapter by leaving the last word to Lewis Carroll. As he probably would have pointed out, there is no reason to impose a universal set of definitions of rationality, but there are plenty of reasons why, if we are about to claim that animals are or are not rational, we should make a serious effort to define what we mean.

> I maintain that any writer of a book is fully authorised in attaching any meaning he likes to any word or phrase he intends to use. If I find an author saying, at the beginning of his book "Let it be understood that by the word 'black' I shall always mean 'white' and by the word 'white' I shall always mean 'black'," I meekly accept his ruling, however injudicious I may think it. (Carroll 1964/2000, p. 226).

Acknowledgments

I am deeply grateful to the Editors for the very substantial comments and suggestions improving previous versions of this paper. Thanks also to Martha Klein and Marian Dawkins for thoughtful and useful advice. This paper was supported with financial support from the BBSRC (Grant 43/S13483) and benefited from a partial research leave spent at the Institute for Advanced Studies in Berlin.

References

Arrow, K. J. (1986). "Rationality of self and others in an economic system." Journal of Business **59: S385-S399**.

Bateson, M. (2002). "Context-dependent foraging choices in risk-sensitive starlings." Animal Behaviour **64**: 251-260.

Bateson, M., S. D. Healy, and A. Hurly. (2002). "Irrational choices in hummingbird foraging behaviour." Animal Behaviour **63**: 587-596.

Bautista, L. M., J. Tinbergen, and A.Kacelnik. (2001). "To walk or to fly: how birds choose among foraging modes." Proceedings of the National Academy of Sciences **98**: 1089-1094.

Becker, G. S. and R. J. Barro (1988). "A reformulation of the economic theory of fertility." The Quarterly Journal of Economics **103**: 1-25.

Brown, H. I. (1995). Rationality. The Oxford Companion to Philosophy. T. Honderich. Oxford, Oxford University Press**:** 744:745.

Carroll, L. (1964/2000). Symbolic Logic and the game of logic. The Annotated Alice. M. Gardner. London, Penguin.

Chappell, J. and A. Kacelnik (in press). "Selection of tool diameter by New Caledonian crows *Corvus moneduloides*." Animal Cognition **online first**.

Cosmides, L. and J. Tooby (1997). Dissecting the computational architecture of social inference mechanisms. Characterizing human psychological adaptations. G. Bock and G. Cardew. Chichester., Wiley. **208:** 132-158.

de Groot, A. D. (1965). Thought and Choice in Chess. The Hague, Mouton.

Fehr, E. (2003). "The nature of human altruism." Nature **425**: 785-791.

Fernández-Juricic, E., S. Siller, and A. Kacelnik. (In press). "Flock density and social foraging in starlings: an experimental analysis of scanning behavior." Behavioral Ecology.

Gigerenzer, G., P. M. Todd, and the ABC Research Group. (1999). Simple heuristics that make us smart. New York, Oxford University Press.

Giraldeau, L.-A. and T. Caraco (2000). Social Foraging Theory. Princeton, Princeton University Press.

Grafen, A. (1999). "Formal Darwinism, the individual-as-maximising-agent analogy and bet-hedging." Proceedings of the Royal Society of London B **266**: 799-803.

Grafen, A. (2000). "A biological approach to economics through fertility." Economics Letters **66**: 241-248.

Hamilton, W. D. (1964). "The genetical evolution of social behaviour." J. Theor. Biol. **7**: 1-52.

Houston, A. I. (1997). "Natural selection and context-dependent values." Proceedings of the Royal Society of London B **264**: 1539-1541.

Houston, A. I., A. Kacelnik, and J, McNamara. (1982). Some learning rules for acquiring information. Functional Ontogeny. D. J. McFarland. Boston, Pitman**:** 140-191.

Hunt, G. R. (1996). "Manufacture and use of hook-tools by New Caledonian crows." Nature **379**(6562): 249-251.

Hunt, G. R. (2000). "Human-like, population-level specialization in the manufacture of pandanus tools by New Caledonian crows *Corvus moneduloides*." Proceedings of the Royal Society of London B **267**(1441): 403-413.

Hunt, G. R. and R. D. Gray (2004). "The crafting of hook tools by wild New Caledonian crows." Proceedings of the Royal Society of London B (Suppl.) **271**(S3): S88-S90.

Hurly, T. A. and M. D. Oseen (1999). "Context-dependent, risk-sensitive foraging preferences in wild rufous hummingbirds." Animal Behaviour **58**: 59-66.

Kacelnik, A. and M. Bateson (1996). "Risky theories - the effects of variance on foraging decisions." American Zoologist **36**: 402-434.

Kacelnik, A. and I. C. Cuthill (1987). Starlings and optimal foraging theory: modelling in a fractal world. Foraging Theory. A. C. Kamil, J. R. Krebs and H. R. Pulliam. New York, Plenum Press**:** 303-333.

Krebs, J. R. and A. Kacelnik (1991). Decision Making. Behavioural Ecology: An Evolutionary Approach. J. R. Krebs and N. B. Davies. Oxford, Blackwell Scientific Publications**:** 105-136.

Krebs, J. R. and R. H. McCleery (1984). Optimization in behavioural ecology. Behavioural Ecology: and evolutionary approach. J. R. Krebs and N. B. Davies. Massachussets, Sinauer: 91-121.

Luce, R. D. (1977). "The choice axiom after twenty years." Journal of Mathematical Psychology **15**(3): 215-233.

Mas-Colell, A., M. D. Whinston, and J.R. Green. (1995). Microeconomic Theory. Oxford, OUP.

Oaksford, M. and N. Chater (1998). An Introduction to rational models of cognition. Rational Models of Cognition. M. Oaksford and N. Chater. Oxford, OUP.

Pepperberg, I. M. (1999). The Alex studies: Cognitive and communicative abilities of grey parrots. Cambridge, MA, Harvard University Press.

Povinelli, D. J. (2000). Folk Physics for Apes. Oxford, UK, Oxford University Press.

Powell, R. W. and W. Kelly (1977). "Tool use in captive crows." Bulletin of the Psychonomic Society **10**(6): 481-483.

Schuck-Paim, C. (2003). The starling as a rational decision-maker. Zoology Department. Oxford, University of Oxford.

Schuck-Paim, C. and A. Kacelnik (2002). "Rationality in risk-sensitive foraging choices by starlings." Animal Behaviour **64**: 869-879.

Sen, A. (1977). "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory." Philosophy and Publich Affairs **6**(4): 317-344.

Shafir, S. (1994). "Intransitivity of preferences in honey-bees - support for comparative-evaluation of foraging options." Animal Behaviour **48**(1): 55-67.

Shafir, S., T. A. Waite, and B.H. Smith(2002). "Context-dependent violations of rational choice in honeybees (*Apis melifera*) and gray jays (*Perisoreus canadensis*)." Behavioural Ecology and Sociobiology **51**: 180-187.

Simon, H. A. and J. Schaeffer (1992). The Game of Chess. Handbook of game theory. R. J. Aumann and S. Hart. Amsterdam, Elsevier. **1:** 1-17.

Simonson, I. and A. Tversky (1992). "Choice in context: trade-off contrast and extremeness aversion." Journal of Marketing Research **29**: 281-295.

Tomasello, M., J. Call, and B. Hare (2003). "Chimpanzees understand psychological states - the question is which ones and to what extent." Trends in Cognitive Sciences **7**(4): 153-156.

Tversky, A. and I. Simonson (1993). "Context-dependent preferences." Management Science **39**(10): 1179-1189.

Vásquez, R. A. and A. Kacelnik (2000). "Foraging rate versus sociality in the starling Sturnus vulgaris." Proc. Roy. Soc. Lond. B. **267**: 157-164.

Visalberghi, E. (2000). Tool use behavior and the understanding of causality in primates. Comparer ou prédire: Examples de recherches comparatives en psychologie aujourd'hui. E. Thommen and H. Kilcher. Fribourg, Suisse, Les Editions Universitaires**:** 17-35.

Waite, T. A. (2001). "Background context and decision making in hoarding gray jays." Behavioral Ecology **12**(3): 318-324.

Waite, T. A. (2001). "Intransitive preferences in hoarding gray jays (*Perisoreus canadensis*)." Behavioural Ecology and Sociobiology **50**: 116-121.

Weir, A. A. S., J. Chappell, and A. Kacelnik. (2002). "Shaping of hooks in New Caledonian crows." Science **297**(5583): 981.

Figures and labels

Figure 2.1

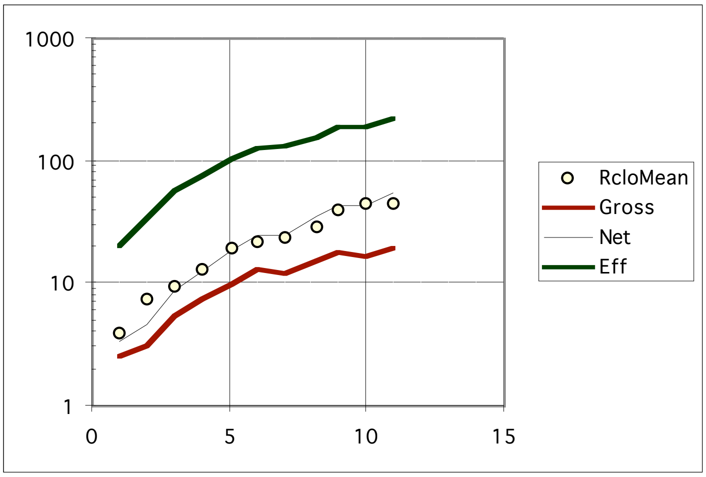

Figure 2.1 label:

To walk or to fly? Starlings were offered a choice between obtaining a food reward by a number of walks or a number of flights. The number of fights (abscisa) was experimentally fixed, and the number of walks at indifference (ordinate) was obtained by titration. The symbols show the average indifference point among four starlings. The lines, from top to bottom, show the predictions of choice according to maximum gain per unit of expenditure, maximum net gain per unit of time and maximum gross gain per unit of time. The agreement with the middle line indicates that starlings make these choices in a way that maximises net rate of energy gain. Modified from (Bautista, Tinbergen et al. 2001)