Animal Rationality:

Are any nonhuman animals rational? What issues are we raising when we ask this question? Are there different kinds or levels of rationality, some of which fall short of full human rationality? Should any behaviour by nonhuman animals be regarded as rational? What kinds of tasks can animals successfully perform? What kinds of processes control their performance at these tasks, and do they count as rational processes? Is it useful or theoretically justified to raise questions about the rationality of animals at all? Should we be interested in whether they are rational? Why does it matter?

Section 1.1 of this chapter provides a landscape of theoretical issues and distinctions that bear on attributions of rationality to animals; it can be read independently of the synopses of chapters, which follow in the remaining sections. These summarize the arguments of the chapters in each of the volume's six parts, on: types and levels of rationality (section 1.2); rational versus associative processes (section 1.3); metacognition (section 1.4); social behavior and cognition (section 1.5); mind reading and behavior reading (section 1.6); and behavior and cognition in symbolic environments (section 1.7). The introduction as a whole aims to provide a substantive and self-standing survey of the topic of animal rationality that is accessible to students and researchers in different disciplines, including philosophy and various sciences.

1.1. Background: theoretical questions and distinctions.

Before considering more closely what is meant by "rationality", it is helpful to draw some rough contrasts between rationality and certain other traits and capacities and to consider the relationships between them.

1.1.1. Rationality and intelligence. We may be ready to admit that nonhuman animals are intelligent, especially in specific domains, but we tend to regard rationality as a special feature of human beings. Aristotle thought that what was distinctive about the human animal was its rationality. To assess the human claim to monopolize rationality, we need a clearer idea of what the difference is between intelligence and rationality (see the chapters by Herman and by Pepperberg).

Intelligence seems in some ways to be broader: it can be expressed in memory, learning, perceptual, and imaginative abilities. Some might describe dogs as intelligent because they can learn complex obedience or tracking tasks, or say that two-year-old children display their intelligence in learning language. Others might prefer to say that an individual dog or two-year old is intelligent on the basis that its capacities in certain domains are exceptional for its species. Such attributions of intelligence need not carry implications of rationality. Nor need rationality imply a high degree of intelligence; some savants might be regarded as rational but not very intelligent.

Should we distinguish intelligence from rationality in terms of certain kinds of behavioural tendency or capacity as opposed to certain kinds of processes underlying and sustaining behaviour? After all, we might remark, "That behaviour was intelligent, but was it the result

of rational processes?" But while the behaviour/process distinction is very important (and we return to it below), it doesn't coincide with the distinction between intelligence and rationality, since it applies to both. For we could just as well remark, "That behaviour was rational, but was it the result of intelligent processes?"

1.1.2. Rationality, generalization, and decentering. Both intelligence and rationality require some degree of flexibility or ability to generalize from one context to another. But rationality seems to involve a more explicit use or recognition of more abstract similarities or patterns. There are differences of degree and of grain here: compare the ability to transfer an existing foraging technique to a new type of food with the ability to transfer transitive inference from social to non-social contexts—say from social dominance rankings to rankings of foraging opportunities. The latter seems to require a more abstract ability, one that is less context-bound.

Moreover, we are inclined to recognize rationality as opposed to intelligence when the ability to generalize involves decentering from me-here-now, and entertaining alternative possibilities: taking into account the past, future or counterfactual possibilities, other places, different possible actions, and other creatures' perspectives. Whether nonhuman animals have such capacities is discussed in several chapters in this volume.2

1.1.3. Rationality and normativity. Rationality in a sense that contrasts with intelligence also requires what philosophers of mind call normativity: there has to be a difference between doing the right thing and doing the wrong thing.3 Suppose an animal appears to be trying to achieve some goal and to be making an error in relation to that goal. What basis is there for saying that the animal has made a mistake in the way it attempts to achieve its goal, rather than that it wasn't attempting to achieve that goal in the first place? Ironically, rationality requires the possibility that the animal might err. It can't be automatically right, no matter what it does; the possibility of mistakes and errors must make sense. A purely normative theory of rationality has no descriptive ambitions: it simply says what an agent should do, or believe, in order to be rational. But even descriptive theories of rationality—the kind relevant to nonhuman animals, so of concern here—must have a normative aspect: when we say that an agent has acted rationally, we imply that it would have been a mistake in some sense to have acted in certain different ways. 4 It can't be the case that anything the agent might do would count as rational. This is normativity in a quite weak sense.

The requirement of normativity can be interpreted in a stronger, reflective, way, as the demand that a rational animal can in some way recognize that it can make mistakes. On this view, it's not enough for there to be a possibility of mistakes only relative to, say, an evolutionary function that we theorists attribute to some behaviour. Rather, there has to be a possibility of mistake for the animal (see and cf. Dennett 1996, ch. 2). Otherwise, although the animal's behaviour can described as if it were rational or irrational from some external perspective, the animal won't really itself have acted for reasons. The capacity to recognize that one can make mistakes can be viewed as part of what distinguishes rationality from mere adaptedness, mere "as if" rationality.

1.1.4. Rationality and consciousness. If rationality makes this stronger reflective

requirement, of a capacity to recognize the possibility of mistakes, that may suggest a relationship between rationality and consciousness. Is either necessary for the other? Here it is important to distinguish two aspects of consciousness.6

First, there is reflective consciousness, consciousness of or about one's mental states—a kind of higher-order mental state, or access to one's mental states. It is tempting to find links between rationality and reflective consciousness (see Call's chapter herein). It can be argued that recognizing the possibility of mistake, or decentering from self to recognize the possibility of other minds, requires higher-order mental states—mental states about other mental states. If such arguments are right, then reflective consciousness may fulfil this requirement of rationality.

Second, there is phenomenal consciousness: "what it is like" to be in a given conscious state, say, of seeing red, or experiencing hunger. How is phenomenal consciousness related to rationality? Is the presence of information to phenomenal consciousness either necessary or sufficient for it to be available as a reason for acting? Arguably, if information is not available to the animal as a reason for acting then it is not present to phenomenal consciousness either. But could information provide an animal's reason for acting even if it is not phenomenal-ly conscious? It could be argued that a sophisticated robot might in principle be a rational agent even if it lacked phenomenal consciousness. However, some theories try to explain phenomenal consciousness in terms of higher-order mental states, in effect linking phenomenal to reflective consciousness. If they are right, a role for reflective consciousness in rationality may also give phenomenal consciousness a role in rationality.

1.1.5. Rationality and conceptual and linguistic abilities. We began with a contrast between rationality and intelligence. Two features–the capacity to generalize in a relatively abstract way including to decentre from me-here-now, and the capacity to make (and perhaps to recognize that one can make) mistakes–appear to be central to what is distinctive about rationality. If this is roughly right, it immediately raises several further is-sues about the relationship of rationality to conceptual abilities, language, and reasoning. Must an animal have concepts and conceptual abilities in order to generalize and recognize mistakes? Indeed, must it have language?

For example, can an animal decentre from itself and take account of the perspective of another animal by simulation–by using its own practical abilities off-line–or does such decentring require theorizing about the other's mind and so require having both concepts of and a theory about mental states? Can simulation underwrite a sense in which animals can be rational despite lacking conceptual and theoretical abilities?8 Must an animal have the concept of a mistake or of a reason in order to recognize that it has made or is likely to make a mistake--or could this again be registered and manifested in practical terms?9 In the absence of the fine-grained distinctions that language makes available, how can we design experiments to assess such capacities for decentering or recognizing mistakes?10 What kinds of nonverbal behaviour, under what conditions, might provide evidence for these capacities? How might symbols support or enable rationality?11

1.1.6. Rationality and reasoning; behavioural versus process rationality. The above

questions about the roles of reflective consciousness and of conceptual and linguistic abilities in rationality in turn point to an underlying issue about the relationship between rationality and reasoning. Reasoning is a special kind of behaviour-generating process, which might be thought to involve reflective consciousness, or conceptual or linguistic abilities. Must rational behaviour be generated by reasoning, or, more generally, by a rational process? Is rationality a feature of behavioural patterns and capacities, or of the processes that sustain behaviour? This latter question is of central importance in considering how and in what ways animals may be rational.

An animal's behaviour can be described as rational from an external perspective without implying that the processes that generate that behaviour are rational. For example, behaviour might be described as rational because it maximizes utility, or increases genetic fitness (see Kacelnik's discussions of E-rationality and B-rationality in this volume). Any animal, if it is to satisfy its goals and to survive and reproduce, must act in the world: to do the appropriate thing, given its circumstances, and to adjust its behaviour as its circumstances change. We can ask whether an animal's behavior is rational in this sense.

But an animal is rational in a different sense if its behaviour is rationally selected by the animal because it is appropriate to its circumstances. There is a sense in the genetic fitness of certain behaviour may explain it; but it doesn't give a rational explanation of the behaviour: fitness doesn't show that the processes that produced the behaviour were rational. When we ask whether animals are rational, we are interested not just in whether their behaviour is rational, but also in whether rational processes explain their behaviour. That is, we want to know not just whether its behaviour is rational, but if so, why that is so: we want to know how the animal selects behaviour appropriate to its circumstances. In particular, we want to know whether rational processes explain how an animal manages to behave rationally.

In application to human beings, different disciplines use "rationality" in different senses; some focus on rational behaviour and others on rational processes, in line with their different assumptions and purposes (as Kacelnik's chapter further elaborates). As a result, there is a danger of talking at cross-purposes in interdisciplinary discussion. Some further background may help to avoid this danger by showing how the distinction between behavioural rationality and process rationality arises in various contexts.

1.1.7. Ends versus means. The rationality of ends can be distinguished from the rationality of means, or instrumental rationality. Instrumental rationality is concerned with finding the best means to a given end or goal–something we desire or need or want–rather than with the selection of ends.12 It concerns whether, given certain ends, an agent is appropriately sensitive to information about means/ends contingencies.13 "Appropriate sensitivity" can be understood in different ways, in terms of behaviour or the processes that generate it (see below). An important question here is whether the sensitivity to means/ends information appropriate to instrumental rationality requires, or is enhanced by, representations of causality (see Call's chapter) or of the intentional states of others (see the chapter by Clayton et al, and those in Part V).

The rationality of ends themselves is more problematic to assess. Do we have any basis for

holding certain human goals to be irrational? (We'll return to this question below, in discussing the relationships between formal and substantive rationality.) In the animal case, if we appeal to biological functions to assess an animal's goals, are we addressing their rationality at all, as opposed to their adaptiveness?

1.1.8. Instrumental rationality as behavioural rationality. Instrumental rationality can be understood as either behavioural or process rationality; classical conceptions of human rationality in different disciplines contrast in this respect. Instrumental rationality is classically conceived by economists in terms of consistent patterns of behaviour (what Kacelnik calls "E-rationality") and by philosophers and many psychologists in terms of the processes that generate behaviour (what Kacelnik calls "PP-rationality").

A reasoning process may be reflected in the theory of why the behaviour it would lead to counts as rational. But behavioural rationality does not itself require that reasoning be rehearsed by the agent or that any particular process actually generate her behaviour. Behavioural rationality is thus in principle open to the possibility that the processes that actually generate an agent's rational behaviour do not correspond to the theoretical account of why the behaviour counts as rational. In this sense, behavioural rationality is in principle compatible with the use of rules of thumb or heuristics. Heuristic are decision-making processes that can reliably generate rational behaviour in specific contexts or environments, even though they do not implement ideal reasoning processes and so may not generate rational behaviour in other contexts or environments.

A further question is just how reliable the use of heuristics must be, across what range of variation in possible environments, in order to be compatible with rational behaviour. Some perspective on this question can be provided by a potted four-stage history of conceptions of behavioural rationality.

(a) Classical behavioural rationality. In economics, idealized normative theories of behavioural rationality are exemplified by orthodox expected utility theory (EUT), on which orthodox game theory builds (see below for more on the difference between individual rationality, as in utility theory, and strategic social rationality, as in game theory). EUT says that the rational choice from a range of alternatives is the one that maximizes the agent's expected utility, that is, the sum of the probability-weighted utilities of each possible outcome of a possible action. This is a conception of instrumentally rational behaviour: possible choices are assessed in terms of their likely consequences, as means to the agent's ends. Alternative means by which an agent's ends might be pursued are assessed in terms of the degree to which their consequences are likely to fulfil the agent's ends. EUT is infinitely flexible, and domain-general rather than domain-specific. That is, it is not tied to environments with any particular structure, but should in principle work equally well in any environment (if–a big "if"–we abstract from the costs of applying it). Regardless of whether such classical theories of rationality describe the actual processes that generate behaviour, they were regarded as broadly predictive of human behaviour—though of course lapses here and there were admitted.

(b) Heuristics and biases. A wave of debunking experimental research was ushered in by

Kahneman and Tversky's seminal work, which showed that the norms and/or predictions of classical theories of behavioural rationality were violated systematically by human decision-makers, in a variety of specific, well-charted ways. Decision-makers' patterns of choice were better described in terms of various heuristics and biases,14 some domain-specific, than by classical conceptions of behavioural rationality such as EUT. These findings were generally taken to be bad news for rationality: to show that human behaviour was less rational than might have been expected or hoped (rather than that rationality had been misunderstood by EUT). From the point of view of behavioural rationality, the bad news was not that much human behaviour is generated by the use of heuristics rather than rational processes; it was the systematic irrationality of the behaviour. Heuristics would have been acceptable had they reliably generated rational behaviour. The problem was that they didn't.15

(c) Ecological rationality. A more positive reaction to the debunking evidence eventually emerged, represented by Gigerenzer's et al's (1999) slogan that "simple heuristics make us smart" and accompanying scepticism about the unrealistic "demonic" character of classical conceptions of rationality, with their disregard for decision-making costs and for the practical limitations on real decision-makers. Further experimental work on ecological rationality demonstrated that in the right environments reliance on simple heuristics, such as the familiarity heuristic, or the one-good-reason heuristic, produces behaviour with better consequences than reli-ance on more sophisticated and flexible but costly domain-general decision-making processes. Evolution, it was argued, won't build the costly flexibility of demonic processes that produce rational behaviour in all possible environments into decision-makers. Rather, it will sensibly fit them out with a toolkit of heuristics adapted to and triggered by interaction with environments they are likely actually to face.16 If relevant envir-onments reliably include certain informational resources, evolution may not duplicate them internally (see and cf. Brooks 1999). On this view, rational behaviour is environmentally situated; the processes that generate it are, in effect, distributed across agents' heuristics and the environments with which agents are interactively coupled. The structure of specific environments and information contained in specific environmental processes does not just contribute to the effects of behaviour, but can contribute to the processes that generate rational behaviour.

(d) Machiavellian social rationality. Recently Kim Sterelny17 has argued that it is no accident that proponents of ecological rationality tend to focus on behaviour in non-social environments. Sterelny explains that natural non-social environments are transparent or at least translucent in relation to information about an agent's most effective means to his ends; by contrast, social environments are often informationally opaque and hostile, as a result of strategic thinking, deception, mimicry, and informational manipulation. Nature doesn't try to predict and outwit an individual agent deciding on the best means to achieve his ends;18 but other agents often do, in order to better achieve their own ends. Evolutionary pressures favouring deceptive mimicry and other ways of manipulating information in social environments contribute significantly to the complexity of social information. As a result, social environments place high demands on cognition. It is implausible that these demands can be met in a way that would generate rational behaviour by simple heuristics that rely on information in environmental structures, given that such information is subject to

manipulation and deception. This position draws on the Machiavellian Intelligence hypothesis19: that social life, with its potential for deception and manipulation of information, drives the development of advanced cognitive capacities, which may include capacities for social cognition and metacognition.

If simple heuristics cannot cope with informationally hostile social environments, what is the next bounce for conceptions of instrumental behavioural rationality? Such conceptions might retreat from ecological optimism and revert to a classical view of normative behavioural rationality after all--though one tempered by the realization that human beings fall systematically short of rationality and so are liable to behave irrationally in challenging social environments to which their simple heuristics are inadequate. Or, is there a plausible way for an ecological view of rationality to go social? Are there social heuristics--heuristics well-adapted to social environments--that reliably generate behaviour fit to achieve one's ends in social environments and that are unlikely to be undermined by deception and the manipulation of information?20 Can the processes that generate rational behaviour be distributed across an agent's social heuristics and her social environment (including other agents), despite the informational hostility of social environments?

1.1.9. Instrumental rationality as process rationality. By contrast with human rationality as understood in economics, human rationality as conceived in philosophy and parts of psychology is process rationality (as in Kacelnik's PP-rationality). Instrumental process rationality is a feature of the way we select the means to achieve our ends. It is not sufficient for process rationality that someone's behaviour is appropriate to his ends, because this may result from coincidence or accident. A rational process must lead to the right results reliably, not accidentally. Behavioural rationality necessarily concerns behaviour and is in that sense practical; but process rationality can concern the processes that lead to beliefs as well as the processes that lead to behaviour. That is, process rationality includes theoretical as well as practical rationality (see below for more on this distinction). In the practical instrumental case, process rationality reliably selects the action that is most likely to achieve an agent's ends; in theoretical case, process rationality leads to beliefs that are true or likely to be true. The essential idea is that of getting things right in the right ways. Behaviour is rational in a sense that derives from the rationality of the process that generates it: rational behaviour is non-accidentally appropriate to my goals, because it results from the right kind of process, a rational one.

When an animal's behaviour on some task is functionally similar to human behaviour on a similar task, it may be tempting to explain the animal's behaviour in the same ways we would explain the human behaviour (see the chapter by Shettleworth and Sutton)—say, in terms of processes of inference and reasoning, or metacognition, or mind reading. But different processes can lead to similar behaviour (even among human beings, for example, in recovery of function after brain damage). Without language, how can we determine whether to attribute a rational process to explain an animal's behaviour? More fundamentally, what kinds of processes are the "right kind" to count as rational?

(a) Classical process rationality: reasoning. A classical conception of process rationality views a rational process as a reasoning process:

If someone acts with an intention then he must have attitudes and beliefs from which, had he been aware of them and had he the time, he could have reasoned that his act was desirable. If we can characterise the reasoning that would serve we will, in effect, have described the logical relations between descriptions of beliefs and desires and the description of the action, when the former gives the reasons with which the latter was perform-ed. We are to imagine, then, that the agent's beliefs and desires provide him with the premises of an argument.

On this view, being rational is a matter of working out the means to one's ends and of reaching decisions and judgements by a normative process of reasoning: by carrying out explicit probabilistic, logical, or decision-theoretic inferences. Both theoretical and practical instrumental rationality are seen to demand reasoning; reasoning can be aimed at solving a problem, making a decision, planning a course of action, arriving at a judgement or prediction. In practical instrumental reasoning we reason about how to achieve our ends; in theoretical reasoning we reason about what is true or probable.

What is characteristic of reasoning, and why is it a good thing to make decisions in this way? Reasoning arguably has the core features of flexible generality and normativity--which, as we've seen, are associated with rationality--in rather strong forms.

Reasoning requires a capacity to generalize and to go beyond me-here-now. For example, means-ends reasoning requires the agent to think about different nonactual actions and their possible results, to think counterfactually. And reasoning arguably makes this requirement of flexible generality in a strong form: Reasoning is an entirely general capacity; its application is not restricted to particular topics or environments, and extends fully counterfactually. If someone can reason about one thing then they can reason about anything they can think about. The domain-generality and content-neutrality of reasoning reflects its logical structure: you get the right answer irrespective of the content of the premises so long as the reasoning is valid. Given an end and some beliefs about the best means to achieve it, reasoning gives the agent a general-purpose way to work out what to do. If her beliefs are correct, and she makes no mistakes in the working out, the resulting action will achieve the end, or be more likely to do so than any alternative.

Reasoning also is normative in the sense that there is a difference between reasoning correctly and making a mistake in reasoning. It arguably requires thinking processes to fall under certain cognitive norms or rules, in the sense that thinking should satisfy these rules, though it may fail to do so. This does not itself require that the agent consciously apply and follow the rules, trying to avoid mistakes; his thinking processes just may follow the rules, fairly reliably. Stronger forms of normativity (as explained earlier) also require a capacity to recognise that one can make mistakes, or reflective consciousness, or even a capacity to justify one's decisions/judgments.

Reasoning is often thought to involve some such stronger form of reflective normativity: to requires not just thinking in accord with rules but also a capacity for reflection on the processes of thinking that lead to a deci-sion, for some form of metarepresentation (see Bermudez 2003, ch. 8 and 9). Not only should it be no accident that one's behaviour is

rational, but it should be no accident that the thinking that leads to one's behaviour is rational. It might be held that normativity in the stronger sense enables reasoners to know that the processes that lead to their decisions are likely to lead to the right results. This is because it requires them not just to select the right decisions in the right way, but also to select the right processes for selecting those decisions in the right (reflective) way. Moreover, the ability to reflect on their thinking processes enables reasoners to correct their mistakes. On the other hand, a strong, reflective requirement of normativity threatens a regress of rules for applying rules for applying rules…(as in the tale of Achilles and the Tortoise; Lewis Carroll 1895). This may mean that the supposed advantages of some stronger requirements normativity are illusory.

Classical reasoning processes, then, are usually thought of as requiring a capacity to generalize in the strong form involving domain-neutrality. They can be thought of as taking weaker or stronger forms in respect of normativity: weaker forms require rule-following and the possibility of mistake, while stronger forms require reflective normativity and/or a capacity to provide justifications. It may be attractive to think of rational processes to require reasoning, though not to require the stronger reflective forms of normativity.

(b) Process rationality without reasoning? Processes with the unrestricted generality and flexibility of reasoning can be costly, as proponents of bounded and ecological rationality argue. Perhaps the costs outweigh the benef-its. Perhaps fully domain-general reasoning is too much to require of rational processes. Can a more inclusive characterization of a rational process be given, which retains the aspect of generality (as well as of normativity) in a weaker for? A number of the contributors to this volume consider how processes other than explicit classical reasoning processes might nevertheless count as rational processes explanatory of animal behaviour.

What kinds of processes other than classical reasoning processes might count as rational, and why? Can the use of heuristics count? Must rational processes at the personal (or animal) level be enabled by subpersonal proces-ses with a corresponding or isomorphic structure, in order genuinely to provide a rational explanation of behav-ior? Does explanation of behaviour in terms of associative mechanisms displace explanation in terms of ration-al processes? Can rational processes or their subpersonal enabling processes be extended–widely distributed across neural, bodily, and situating environmental processes–or must they be contained in the animal?

(c) Use of heuristics. While reasoning is a paradigmatic rational process, it may not be as prevalent as we tend to assume, even among human beings. It might be thought that people often arrive at beliefs and perform actions based on perceptual experiences, in ways that do not involve reasoning (even if they reason about the relations among their beliefs). Moreover, as we've seen, there is evidence that people tend systematically to use heuristics and display biases that are not consistent with domain-general reasoning. While the success of heuristics in generating behaviour that achieves our ends may be restricted to specific domains, if those include the domains that matter most to an agent, then perhaps such success is sufficient for rationality. And successful heuristics can be fast and frugal, so less costly than reasoning. Can rationality be understood in terms of an adaptive toolbox of such

domain-specific heuristics, layered together by evolution so as collectively to cope flexibly with a wide range of relevant problems?22 Might arriving at decisions in accord with such heuristics count as a rational process, even if not a process of reasoning?

There are two ways to understand this suggestion, as involving either a less radical or a more radical departure from classical reasoning. On the less radical view, although people can reason about what to do in the classical way, it is often rational to use heuristics instead in certain contexts to decide what to do, because it is known that they are just as likely to be successful in those contexts, and they are easier to use. In this sense, the use of heuristics may be selected reflectively and on domain-general principles, even if heuristics aren't domain-general. We may use reasoning to select which heuristics to use, even if using heuristics is not itself a kind of reasoning; such metacognition may be enough to make the use of heuristics a rational process, or part of one.

On the more radical view, metacognition is not needed for rational processes; using heuristics can be a rational process even if their use is not in turn selected by a process of reasoning. Perhaps we don't decide reflectively to use heuristics, on the general grounds that they are just as likely to lead to success in certain contexts and are less costly. Rather, using them may just be an evolved feature of the way our minds work, perhaps implemented by associative mechanisms. Such processes generate actions that are likely to achieve our goals reliably (in the right environments), not by accident, even though there is no higher-order reasoning process, reflective and domain-general, that explains why we use heuristics. The more radical suggestion is that the use of heuristics could still qualify as a rational process. It may meet weaker requirements of normativity and generality than does reflective, domain-general reasoning. On this view, classical reasoning is too restrictive as a conception of process rationality even in the human case.

This would make more room for process rationality in animals. Many animals have problem-solving capacities that are specific to a particular task or environment, which they cannot generalize and apply to other tasks. Arguably, domain-general reasoning requires logical concepts and linguistic capacities that animals lack.23 But if domain-general reasoning is too restrictive as a conception of process rationality in the human case, then it shouldn't be required for process rationality in animals either.

(d) Personal level versus subpersonal level processes, and associative processes. Following Dennett (1969, 1978, 1991), many philosophers distinguish descriptions of contentful mental states and processes attributed to persons from subpersonal descriptions of information being processed and passed between subsystems (see Hurley's chapter for a similar distinction for animals). Cognitive science shows us how the mental lives of persons are enabled by subpersonal information processes (McDowell 1994a). Subpersonal processes can be described functionally or in terms of their neural implementations. Some subpersonal processes are described in cognitive terms, involving representations; some in terms of symbolic representations; others merely in terms of associative mechanisms.24

Reasoning is a personal-level process; it seems natural to assume that rational processes more generally are also personal-level processes. But it might be held that a genuinely rational

process, one that provides a rational explanation of behaviour, must be enabled by subpersonal processes that correspond to it structurally in certain ways. For example, the subpersonal cognitive processes enabling rational thought processes might be required to have a structure isomorphic to the structure of rational thought.25 This may be regarded as part of what it is for a rational process to be the right kind of process. It wouldn't be enough, on this view, for subpersonal cognitive processes to enable people reliably to arrive at the right answers, if those subpersonal processes bear no intelligible relationship to personal-level processes.

Related issues also arise for animals. It may be tempting to explain the successful performance of certain complex tasks by animals in terms of cognitive processes such as inference, following abstract rules, or metacognition.26 But often sophisticated performances can also be explained in terms of associative mechanisms (see the chapter by Papineau and Heyes). How are associative explanations related to cognitive explanations (explanations in terms of representations), and to rational explanations "at the animal level"? Does some version of an associative/cognitive distinction hold up? Do associative mechanisms implement explanatory rational processes, or do they undercut them, leaving them with no real explanatory work to do? If associative explanations exclude rational explanations, should we always prefer the former in accord with Morgan's Canon, on the grounds that they are simpler or "lower" processes ?27 What kind of behaviour, if any, firmly resists explanation in associative terms, and does it matter?

(e) Widely distributed processes: extended rationality? It is a familiar idea that representations and cognitive processes may be distributed. This is often assumed to mean: distributed within the head, as in distributed neural networks. However, the distribution may be wider than that: the situated cognition, extended mind, and ecological rationality movements suggest the possibility that the cognitive processes explanatory of behaviour can be distributed across agents and the information-carrying environments with which they interact. This view in turn raises the questions of where rational processes can be located, and how they are bounded. Must rational processes be wholly internal to the rational animal's brain or body, or can they be distributed across the animal's brain, body, and environment? Must the external portions of animal-environment interactions merely stand in causal relationships with rational processes, or can they be part of what constitutes a rational process? On what principles should such a causal/constitutive boundary be drawn? Can we regard an animal itself as rational if its ongoing behaviour is best explained by such extended processes, not just by internal processes? If so, are the boundaries of extended rational processes provided by biological evolution, or by culture? What is the role of interactions with social and symbolic environments in human rationality? Can reasoning processes extend to include discussion and argument, or the use of pen and paper in proofs? When animals are enculturated, raised by and with human beings and trained to use symbols, do they become more rational than they would be "naturally"?28

The explanation of behaviour in terms of widely distributed processes, including processes of language use29, begins to blur the distinction between rational processes and rational behaviour. Extended explanations of behaviour are dynamic explanations of behaviour conceived as extending through time. They explain how ongoing patterns of behaviour are

sustained in terms of the dynamic interactions between brain, body, and environment, as one round of movement in a given environment has feedback effects on internal mechanisms that produce the next round of movement, and so on. Bodily movements and their environmental results are an essential part, on the "embodied, embedded" view, of the processes that explain behaviour. Hence they may be an essential part of the subset of rational processes—at least the basis for insisting that rational processes must be wholly internally constituted is unclear. On this conception, rational processes are on view in the world.

A conception of rational processes that is liberal on all the above issues would admit rational processes other than reasoning processes, which could include the use of domain-specific heuristics, processes implemented by non-isomorphic subpersonal mechanisms including associative mechanisms, and widely distributed processes—if these processes are part of what explains why the person (or other animal) reliably gets the right result. If this is too liberal a view of the processes that could provide a rational explanation, how exactly should it be tightened up, and why? A critical question for animal rationality is how to motivate and characterize a middle ground between a requirement of reflective, domain-general reasoning and what might be viewed as excessive liberality about rational processes.

1.1.10. Formal versus substantive rationality. Let's return now to the idea that ends can be assessed for rationality, as well as means. Some light is shed on this idea, as well as on the distinction between behavioural and process rationality, by considering another distinction: that between formal and substantive rationality.

Formal rationality requires that certain formal patterns of preference or expectation, or transitions between them, be respected. For example, if A is preferred to B and B to C, then C should not be preferred to A. Substantive rationality by contrast requires that preferences or expectations have certain contents, that they be accurate or enlightened or prudent or virtuous. For example, it might be regarded as substantively irrational to desire to take up smoking, even if the pattern of your preferences and expectations was formally consistent. The relationship between formal and substantive rationality has been important for contemporary debates in decision theory. While officially decision theory is concerned with the rules of formal rationality, a problem arises if the contents of preferences and expectations can be endlessly pruned and reinterpreted to make them fit formal requirements.30 For example, perhaps your choices can be regarded as meeting a requirement of formal consistency if your preferences are interpreted to give you the end of avoiding pain on all days of the week other than Thursdays. But the substantive irrationality of such an arbitrary end may argue against this explanation of your behaviour as formally rational, and support instead the view that you are formally irrational. In the absence of any constraints on the substantive contents of preferences or expectations, formal rationality (whether understood as explanatory or normative) rules nothing out and becomes empty. So formal rationality appears to depend implicitly on assumptions about the substantive contents of preferences and expectations. And such assumptions may in effect include assumptions about the rationality or irrationality of certain ends. How such assumptions could be explicitly defended is a further question. In the case of animals, biological fitness will probably feature in any answer.31

The emptiness of a purely formal conception of rationality also puts pressure on a purely behavioural conception of rationality, with no element of process rationality. Our behaviour expresses our intentions and preferences, but it does not wear them on its sleeve. The same observable behaviour could be intentional under quite different descriptions, depending on context, and may express quite different preferences. I may choose the pear rather than the chocolate walnut cake because I am on a diet, because I dislike chocolate, because I am allergic to nuts, because it is more compatible with the wine, etc. The object of my choice is accordingly less independent of the processes that lead to it than a purely behavioural conception of rationality may assume. In order for the theorist of rationality to identify correctly the object of my choice, in order to relate it to the object of my choice in other cases and hence to determine whether my patterns of behaviour are rational or not, she may need to understand something about my ends, about what intentions and preferences my choices express, and hence about the process that generates my choices. In judging a pattern of behaviour to be rational, we are implicitly judging a pattern of preferences with certain contents, which my choices express, to be rational (see Kahneman and Tversky 2000a, xiv-xv). In the case of human beings, we may overlook this point because we can resolve it using language: in the way we specify the options, or by asking people to say what they prefer (though linguistically specified choices are associated with persistent forms of irrationality such as framing effects; see Kahneman and Tversky 2000a, b). In the case of animals, language doesn't come to the rescue. But the point remains that behaviour fits different patterns depending on what we take its object to be, and the processes that generate the behaviour will often be relevant to determining that.

1.1.11. Individual versus social rationality. Decision theory concerns individual rationality: what one individual should do who is playing, in the jargon, "against nature", where nature is not another agent. But the situation is much more complex if we move to game theoretic contexts and to strategic rationality, where the agent is playing against other agents who are in turn playing against him. Here, the agents' decisions are interactive. A basic equilibrium solution concept in game theory is that of "best reply". This is a conception of behavioural rationality: my action is rational if it is the best reply for me to what you do, and yours is rational if it is the best reply for you to what I do.

While game theoretic rationality can thus be understood to demand certain behaviours rather than certain decision processes, in the way characteristic of E-rationality, game theory also raises important questions about how we decide what to do in such interactive situations and the function in them of understanding other minds. What is rational for me to do depends on what I predict you will do, and what is rational for you to do depends on what you predict I will do. If we are both rational, we each need to occupy the rational perspective of the other in order to predict the other's behaviour and hence to decide rationally ourselves what to do. Such game theoretic contexts, and social contexts more generally, are a source of pressure to acquire the ability to decentre from self and capacities for social cognition and social rationality: to understand that another agent may have goals and expectations different from one's own, and to use that understanding to predict her behaviour and to perhaps to manipulate her informational environment.32

The study of social interactions and cognition in animals is a topic of great current interest.

What kinds of understanding of one another do their social interactions involve? Can apparently rational social learning be explained in terms of lower-level processes (see the chapter by Addessi and Visalberghi)? In particular, do nonhuman animals attribute mental states to other animals: are they mind readers or are they just very smart behaviour readers? What exactly is the difference between genuine mind reading and mere behaviour reading? How can this difference be assessed; what evidence would justify attributing mind reading to non-linguistic animals?33 Could there be a difference in the functions of mind reading and of behaviour reading, even if the evidence for mind reading is behavioural? Is it helpful to distinguish different kinds or levels of understanding of other minds? There is evidence that children come to understand the desires of others before they understand their beliefs (Rakoczy, et al, submitted). Might animals who fail nonverbal versions of tests for the ability to attribute false beliefs to others nevertheless understand the cognative states of others and pass other, less demanding tests of mind reading? Is whatever social rationality of which animals are capable limited to certain domains or activities? Is it best understood in terms of their possession of a rudimentary theory of mind, or rather in terms of a practical ability to simulate the decision-making processes of other agents?

1.1.12. Practical versus theoretical rationality. The distinction between practical and theoretical rationality, the rationality of actions versus the rationality of beliefs, is important in assessing the rationality of nonhuman animals. Theoretical rationality involves evaluating the evidence for and against beliefs, and arguably depends on conceptual or even linguistic abilities. Practical rationality may be less demanding, and be more plausible to attribute to animals.34 Even if an animal cannot weigh up evidence with due scepticism or assess explanations in a theoretically rational manner, its actions might nevertheless be instrumentally rational, that is, rationally sensitive to information about means/ends contingencies.

The theoretical/practical distinction is relevant to assessing the social rationality of animals. It may be that nonhuman animals have an ability to understand one another that is more practical than theoretical, in two senses: animals may understand the goals and intentions of others more readily than their beliefs, and they may do so by off-line simulation of practical decision-making capacities rather than by theoretical reasoning. Simulation is another candidate for a rational process that is not a classical reasoning process (though simulation can also be understood as subpersonal enabling process).

The theoretical/practical distinction is also relevant to the idea of rational processes that are not fully domain-general. It is natural, if you think of rationality in theoretical terms, to construe it as a domain-general capacity, not restricted to any specific contexts, such as food gathering or social activities. However, there is evidence that even human rationality may be tuned to specific domains, such as the detection of cheating in social contract situations.35 If rationality is understood in practical terms, this domain-specificity becomes more understandable, as practical know-how is often implicit and bound to specific procedures and domains. Moreover, what a creature can do on-line may limit what it can do off-line, or simulate doing. So, if the mind reading abilities of animals are understood in terms of a practical ability to simulate the decision-making processes of others, then we should not be surprised to find that their mind reading skills are limited to specific contexts (see the chapter

by Hurley).

1.1.13. The landscape of questions about animal rationality. So far, we've sketched a landscape of distinctions and issues against which animal rationality can be assessed. To summarize:

* Rationality differs from intelligence
  o in requiring a greater capacity for flexible generalization–extending to some capacity for decentering from me here and now, if not to full domain-generality–and
  o in its normativity, which requires the possibility of making mistakes and perhaps some kind of reflective awareness, or metacognition, concerning the possibility of mistake.
* Disciplines vary in their focus on rational behaviour as opposed to on rational processes that explain behaviour.
* Rational behaviour is usually understood in terms of instrumental practical rationality, concerning the rational means to a given end, rather than the rationality of ends. However, it is arguable that the implicit dependence of purely formal conceptions of rationality on some substantive assumptions puts pressure on behavioural rationality to include some process elements.
* Classical conceptions of behavioural rationality, such as EUT, have been challenged by results from the heuristics and biases line of research, and revised and liberalized by the ecological rationality school.
* Process rationality can be understood either as practical, concerning the processes that explain actions, or as theoretical, concerning the processes than explain beliefs.
* A classical view of process rationality in terms of reflective, domain-general reasoning might be revised and liberalized by allowing nonreasoning processes to count as rational, such as domain-specific heuristics, processes implemented by associative mechanisms, and widely distributed processes.
* Such liberal revisions of classical conceptions of human behavioural and process rationality may make more room for animal rationality.
* In social environments, agents play against other agents (rather than nature), who can manipulate information; this creates selection pressure that may drive the development of advanced cognitive capacities, including behaviour reading and mind reading capacities, in social animals.
* Practical rationality may be more accommodating than theoretical rationality of the forms of social cognition found in nonhuman animals, and of the domain-specificity of many of their problem-solving abilities.

With this background, we'll now survey the chapters. They are divided into parts that focus on: types and levels of rationality; the distinction between rational and associative processes; metacognition; social behaviour and cognition; mind reading and behaviour reading; and behaviour and cognition in symbolic environments.

1.2. Types and levels of rationality.

The contributors in this section explore the relations between behaviour and the processes that explain behaviour, and the senses in which animal behaviour might be rational in virtue

of features other than classical reasoning processes on the human model.

1.2.1. Kacelnik. When different disciplines use a concept in different ways and for different purposes, there is always a danger that they will talk past one another. Zoologist Alex Kacelnik (chapter 2) helps to avert this danger by distinguishing three conceptions of rationality, employed in philosophy and psychology, in economics, and in biology. PP-rationality is typically adopted by philosophers and cognitive psychologists. On this conception, rationality is exhibited when beliefs or actions are adopted on the basis of appropriate reasons. PP-rationality focuses on the process by which belief or action is arrived at, rather than the outcome of beliefs or actions—by contrast with rationality understood as utility maximization, which is a matter of behaviour appropriate to goals. However, beliefs or actions produced in appropriate ways should over time produce appropriate outcomes.

Since PP-rationality is understood in terms of thought processes, Kacelnik suggests that it is not an appropriate conception of rationality with which to address questions about the rationality of animals. From a biologist's perspective, the problem with discovering whether non-linguistic animals are PP-rational is that of finding a way to ascertain what the subject's thoughts are and whether they result from an appropriate process of reasoning. The concepts used in characterising PP-rationality – belief, thought, reasoning, and so on – are closely tied to language and don't have immediate application to non-linguistic animals. But a more general worry is that the processes that produce behaviour are not observable. And even if some animal behaviour may be best explained by reference to representations that aren't directly observable, the causation of behaviour by representations doesn't entail that the behaviour results from the subject's reasoning. This difficulty is not limited to animals, but applies in the human case also. It is hard to tell whether a human subject's belief has been arrived at on the basis of appropriate reasons: "the elements that entered into her reasoning process may [although they produce an appropriate belief] have been influenced by the kinds of mechanism that the present definition would explicitly exclude from rationality."

One response to these problems with the conception of PP-rationality is to switch to a conception of rationality that focuses on outcomes rather than the processes that produce them. There is a conception of rationality prev-alent in economics according to which behaviour that maximises expected utility is rational, no matter how it was produced or selected; Kacelnik calls this E-rationality. In this sense it describes a pattern in behaviour.

A problem with this approach is that the utility to be maximised is understood as whatever is in fact maximised by the observed behaviour of the agent; maximizing behaviour is restricted only by formal requirements of consistency. But many peculiar and perverse forms of behaviour can be interpreted as maximizing some common currency, and hence can be regarded as rational. The question of whether an animal's behaviour is E-rational is thus in danger of trivialization, in the absence of some further constraints on what behaviour counts as rational.

Kacelnik suggests that, rather than assessing patterns of behaviour purely in terms of utility-maximization, we can turn to evolutionary biology to provide more substantive constraints

on maximization. Evolutionary change results, in part, from natural selection. Natural selection produces phenotypic properties that can be predicted using principles of maximisation applied to a defined currency: fitness. Fitness is to biological theory what utility is to economics. But unlike utility, fitness is not revealed by the behaviour of agents; rather, it is defined by the genetic theory of natural selection. The fitness of a biological agent is its degree of success relative to that of other agents in the same population. Genetic material guides the development and functioning of indivi-duals; their behaviour is driven by mechanisms that evolved because they produce biologically rational behav-iour. A B-rational individual is one whose behaviour maximizes its inclusive fitness across a set of evolution-arily relevant circumstances.36 In its emphasis on outcomes rather than processes and on consistency, B-rationality is similar to E-rationality (and indeed there is a tradition of fruitful interactions between evolutionary game theory and economics). But B-rationality is far more constrained than E-rationality in its account of what is to be maximised. As Kacelnik explains, a consistent maximizer of fitness is a consistent maximizer of something, and a consistent maximizer of anything counts E-rational. So B-rationality entails E-rationality in relevant circumstances, but goes beyond it to specify that what is maximized must be inclusive fitness.

Two theoretical weaknesses of this conception are its under-specification of what counts as an evolutionarily "relevant" circumstance and its assumption (rather than explanation) of the set of behavioural options from which the fitness-maximizing behaviour is chosen. Both weaknesses limit its explanatory power.37 Behaviour may not be B-rational in a domain that has not been encountered often enough to influence natural selection – but how often is enough? – or because the animal has not got an adaptive behaviour it its option set—but why hasn't it? Note that lack of a specification of relevant domains can be a problem for B-rationality even if domain-specificity itself is not problematic.

Kacelnik goes on to illustrate how these different conceptions of rationality apply to birds, giving examples of choices of foraging techniques by starlings, choices of flowers by hummingbirds, and the obtaining of inaccessible food by a crow, who makes a tool with which to retrieve it.

B-rationality predicts behaviour in evolutionarily relevant circumstances from models that show which of a limited set of behavioural options would maximize fitness (though models may use various proxies for fitness). For example, an explanation of starlings' choices to forage by walking or by flying assumes that starlings are B-rational, but makes no assumptions about whether starlings use reasoning to make their choices, or about what other mechanisms may explain their behaviour.

In another example, Kacelnik shows how apparent violations of E-rationality and hence of B-rationality by hummingbirds in choosing flowers can be understood as rational in these senses after all. Hummingbirds show context-dependent reversals of choice that are prima-facie inconsistent. A biologist might respond by claiming that one of the contexts was not evolutionarily "relevant", or by reinterpreting the context-dependent choices as responses to different problems so avoiding any inconsistency. The availability of such responses to apparent violations of B-rationality may raise concern about to how robustly inclusive fitness

maximization constrains explanations that assume behaviour is B-rational. Is it still too easy to accommodate apparent violations?

A more general concern is whether, by excluding any element of process rationality, B-rationality tells us any more about behaviour than simply that it is adaptive. It leaves untouched questions about how an animal determines which behaviour is appropriate to its circumstances. How do Kacelnik's foraging birds work out which behaviour satisfies B-rationality? Is this process one of rational decision-making? Is there really nothing useful to be said on these matters?

Kacelnik's final example tempers his pessimism about the applicability of PP-rationality to animals. Betty, a New Caledonian crow, bent a straight piece of wire into a hook, allowing her to retrieve otherwise inaccessible food from the bottom of a vertical plastic well. To do so, she had to leave the site of the problem and find a suitable crack to hold the wire's tip while she bent it. Kacelnik finds it hard to account for Betty's behaviour without attributing to her the capacities to plan ahead, to represent the problem and its solution, and to choose rationally among possible actions. Exercising such capacities seems to amount to something not unlike thinking and perhaps approximates the conditions for PP-rationality. How domain-general these capacities may be in crows is unknown, though Kacelnik doubts that even human reasoning is fully domain-general.

The notion of B-rationality is clearly relevant to many explanatory projects in biology, even though it does not illuminate the processes that enable animals to be B-rational. Many of the contributors to this volume are interested in that further question and in whether those processes can properly be characterised as rational processes. If a positive answer to that question requires animals to engage in reasoning, to have beliefs, desires and other propositional attitudes with conceptual content, or to have linguistic abilities, we may well be sceptical. But there may be other, less demanding ways of understanding the processes that explain animal behaviour as rational.

1.2.2. Dretske. Philosopher Fred Dretske (chapter 3) provides an account of what he calls minimal rationality, which is a feature of something that is done for a reason. There is a difference between something's being caused by an event with a certain content and being explained by the fact that the event has that particular content. Minimal rationality requires not only that behaviour be under the causal control of thought or another representational state, but that it be explained by the content of that representational state. What is done for a reason in this sense need not be done for good reasons, or be the product of reasoning, so minimal rationality is not a normative conception in these senses. Minimal rationality cuts across biological rationality; neither entails the other. It is very close to what Clayton, Dickinson and Emery call "intentional" explanation, in terms of the content of psychological states (see chapter 9).

This account of minimal rationality provides a sense of process rationality that avoids some of the difficulties with PP-rationality; we don't need to appeal to language or to reasoning to characterize a minimally rational process. But it faces the large question of intentional causation. Representational states are intentional states: they are about something. When it

is right to explain a piece of behaviour in terms of what internal representations are about, their content? When does the content of a representation explain its causal function, as opposed to its causal function explaining its content? Given that representations are realised by physical states, when is it possible to explain the causal role of the representational state by appeal to the physical properties of its realisation, and when is it necessary to explain its causal role in terms of its content?

Dretske presents his views by means of examples. While thermostats and plants may have inner states that in a sense represent something, and these states cause certain changes or "behavior", those changes aren't explained by what those states represent. They therefore do not meet the conditions for minimal rationality. For example, what explains why a thermostat closes the electrical circuit to the furnace, thus turning on the heat, is the degree of curvature of a metallic strip with certain chemical properties, not what this represents about temperature. And what explains why the thermostat has this representational state in this causal role in the thermostat is not what it represents, but the intentions of the thermostat's designers, who chose to wire it the way they did: they wanted something that functions the way the metallic strip functions. In another of Dretske's examples, a plant, the Scarlet Gilia, changes from red to white when hummingbirds (who prefer red flowers) depart and hawk moths (who prefer white) arrive. Some chemical state of the plant presumably represents the external conditions in which this color change becomes adaptive for the plant. But what explains the plant's color change is its chemical state, not what this state represents about external conditions. And similarly, what explains why the plant has that chemical state in this causal role is not what it represents, but a selection process operating on the states of the plant's remote ancestor's: natural selection *wanted* something that would function this way.

By contrast, learned behavior in animals can be minimally rational. A foraging bird, for example, may avoid eating a harmless butterfly because it looks like a butterfly that the bird has learned is bad tasting. Some internal state of the bird represents that the bad-tasting sort of butterfly is present, so it avoids it. The bird's behaviour is explained by the content of what it "thinks" about the butterfly. The bird's internal state, like the thermostat's and the plant's, represents something as well as having a causal role. But the content of the bird's representational state, unlike those of the thermostat and the plant, explains why that state has the causal role it has in the bird's behaviour. As a result of a learning process, the bird "thinks" the butterfly tastes bad, and it avoids it because it thinks that. We can't explain why the bird's internal state has the causal role it has – as an element in a process that produces bug-avoidance behaviour – except in terms of what it represents, as a result of learning. Its behaviour thus counts as minimally rational even if it results from associational learning processes.38 Other examples of what Dretske has in mind as minimal rationality are provided by the birds described by Clayton, Dickinson and Emery (chapter 9) and by Herman's dolphins (see chapter 20).

Dretske's distinction between being caused by a state with representational content and being explained by the content of that representational state in effect distinguishes merely mechanical processes from intentional or cognitive causal processes (see the chapter by Clayton, Dickinson and Emery for more on this distinction). This is an important distinction, and provides a sense of "rational process" that does not require reasoning. But it

nevertheless makes very minimal demands of rational processes. We may still want to capture a more robust sense in which processes other than reasoning processes can be rational, for which minimal rationality is a nece-ssary but not sufficient condition. There might be a state in the human visual system that, in virtue of its content, plays a role in explaining our ability to recognise different tools; the process of which that state is a part might be minimally rational, in Dretske's sense. Although we might regard such processes as cognitive – as essentially involving representations – we wouldn't think of them as rational processes in a more robust sense.

1.2.3. Millikan. Philosopher Ruth Millikan (chapter 4) focuses on rational processes that involve what she calls "trials and errors in the head". Evolution conducts trial and error explorations over generations; instrumental conditioning improves on this by allowing an animal to learn within its own lifetime from trials and errors in its behaviour. A capacity to conduct trials and errors in the head goes one better: it is quicker and safer than instrumental conditioning or natural selection. Karl Popper noted the benefits of letting one's hypotheses die in one's stead; Popperian animals, as Daniel Dennett (1996) calls them, can do just that, by trying different possible solutions to problems out in their heads. On one reasonable interpretation of rationality, to be rational is to be a Popperian animal, with the capacity for mental trials and errors. Millikan goes on to distinguish two forms of trial and error in the head, a perceptual/practical form that she thinks is plausibly attributed to some non-human animals, and a more theoretical form that she thinks is not.

Millikan suggests that perceptual representations she calls "pushmi-pullyu representations" can function in such Popperian processes. These are representations that at once describe the animal's environment and direct its behaviour. They are suitable for immediately guiding action because they represent affordances in the animal's environment: they tell the animal where there are things to be picked up, climbed on, eaten, and so on. Millikan suggests that nonhuman animals can engage in trials and errors in the head at the level of perceptions of affordances; this would be a kind of practical rationality. She describes a squirrel that she observed solve the problem of how to reach a bird feeder. The squirrel surveyed the situation from various positions and angles before jumping in a way that required a precisely aimed ricochet to reach the feeder. Millikan suggests that the squirrel may have been engaged in a kind of trail and error in perception. It was trying to see a way to reach the feeder, to see an affordance.39 It was not engaging in syllogistic practical inference; but if rationality is the capacity to make trials and errors in one's head, then its behaviour provides evidence of a form of practical rationality, one shared with human beings. While such behaviour is suggestive, the presence of this rational trial and error process in nonhuman animals would ultimately be settled, Millikan suggests, by neurology and cognitive science – by describing the mechanisms which subserve it. Note that this perceptual/practical form of trial and error in the head is a form of simulation. The squirrel seems to be able to use his perceptual/practical skills off-line to solve a problem, in this case to predict the results of alternative possible actions and select the one most likely to succeed.40

The theoretical form of trial and error in the head may be specific to humans. Non-human animals represent primarily affordances and features of the environment with direct practical significance for the animals: the presence of a hazelnut, or of danger, the location of cached

food, and so on. Their representations of such features are successful when they lead to success in practical activities. But human beings, unlike nonhuman animals, also assiduously collect representations of what Millikan calls "dead facts", which do not pertain directly to practical activity and may not have any practical relevance at all (we never know when a dead fact will come alive). Success in representing dead facts does not turn on practical success, but requires an ability to triangulate different methods of recognizing the same objective state of affairs, and to recognize their consistency or inconsistency. That is, success in representing dead facts requires the capacity to recognize contradictions in thought, which depends on representations with subject-predicate structure that can be internally negated. The capacity to recognize consistency and contradiction is not given in perception. Simple representational systems do not contain contrary representations; for example, multiple signals used to alert con-specifics to danger do not contradict one another. Perceptual representations of affordances don't contradict one another, even if they require choices. Nonhuman animals that primarily represent affordances are unlikely to have this capacity to recognize consistency and contradiction. But it is this capacity that makes possible trials and errors in attempts to represent the world theoretically, in ways decoupled from action, and to test and adjust beliefs until they are consistent through reasoning. Thus, nonhuman animals are unlikely to share with human beings this more theoretical process of trial and error in the head. Their Popperian rationality may be limited to perceptual/practical processes of trial and error in the head.

1.2.4. Bermúdez. A psychological explanation of an animal's action would show why the action made sense from the agent's perspective. In philosopher José Luis Bermúdez' view this would require showing how the actions could result from a process of reasoning. He is thus concerned with how to understand the possibility that non-linguistic animals might be rational in the sense of Kacelnik's PP-rationality. But rather than extend rational processes beyond reasoning processes, he takes them to be correlative, and instead offers an extended account of reasoning that could apply to animals (see also Bermundez 2003).

One way of understanding the process of reasoning involved in PP-rationality, or the theoretical process of trial and error in the head, is in terms of logical operations defined over linguistic structures. As Bermúdez explains, language makes types of reasoning possible that are not possible in its absence, by making it possible to think about thoughts. Such metarepresentation or "intentional ascent" enables both reasoning that exploits the internal structure of a thought and inferences that depend on logical concepts, such as negation or material implication ("if….then…."). But this kind of account can make it very difficult to see how reasoning processes could be present in non-linguistic creatures.

Bermúdez explains how analogues of familiar reasoning processes might be possible in the absence of linguistic structure—that is, how it might be possible for non-linguistic animals to reason without exploiting the internal structure of thought or deploying logical concepts. For example, consider a form of non-linguistic reasoning analogous to this instance of reasoning by modus ponens: If the gazelle is at the watering hole, then the lion is not at the watering hole; the gazelle is at the watering hole; so the lion is not at the watering hole. This would require proto-logical analogues of negation ("not…") and material implication ("if…then…"), which do not require logical concepts, language, or metarepresentation.

Bermúdez explains proto-negation in terms of contrary properties, rather than in terms of a formal operation on whole propositions or thoughts. Proto-negation is available in cases where a creature can think about contrary properties. In Bermúdez' example, the thought that the gazelle is absent from the water-hole is contrary to the thought that the gazelle is present at the water-hole, which allows it to serve a similar role in reasoning to that played by the denial of the thought that the gazelle is present at the water-hole, although it does not result from applying the formal, logical operation of negation. Specific contraries, such as presence and absence, exclude one another, even though they do not do so in virtue of their logical structure. While Bermúdez agrees with Millikan that non-linguistic animals cannot understand the general idea of contrariety, he suggests that they do not need to understand contrariety in order to represent specific contraries (such as presence at versus absence from the water hole) and to reason practically and validly in accord with their contrariety. Note that any resulting reasoning capacities might be expected to be correspondingly domain-specific.

Bermúdez next explains a non-linguistic form of proto-conditional reasoning as a primitive kind of causal reasoning. Conditional reasoning takes one from recognition that there is a conditional dependence between two states of affairs, so that if the first holds, then the second holds, plus recognition that the first state of affairs does indeed hold, to the conclusion that the second holds. Proto-conditional dependence, he suggests, could be represented by non-linguistic animals as causal dependence between events (which may or may not include the agent's own actions, and which may be probabilistic). (While conditionals are often appealed to in analyses of causation, the phylogenetic order of acquisition need not match the order of analysis.) The abilities to represent causal regularities and causal dependencies between events or states of affairs can be of great survival value to animals, and need not involve metarepresentation. Such regularities might include, for example, those between the presence of lions and the absence of gazelles at the watering hole, and between the presence of gazelles and the absence of lions; and the absence of gazelles would depend causally on the presence of lions rather than vice versa. As a result, an animal may be able to reason from a proto-conditional, and employing contraries, as follows: If the gazelle is present at the watering hole, then the lion is absent from the watering hole; the gazelle is present, so the lion is absent.

Such a proto-inference is analogous to modus ponens, but is not valid in virtue of its form in the way logical inferences are, since the transition from premises to conclusion only works because of the specific pair of contraries employed. But despite its domain-specificity, the reasoning is valid in the sense that if the premises are true, then so is the conclusion. And it is clearly of practical value to the animal. Bermúdez concludes that the ability to represent specific contraries—proto-negation–and causal dependencies—proto-conditionals—would make reasoning available to non-linguistic animals, which could in principle underwrite psychological explanations of their behaviour. Note that while Bermúdez and Millikan differ over whether contrariety can plausibly have a role in animal rationality, Bermúdez is primarily concerned to show how forms of proto-inference involving contrariety could be of direct practical relevance to animals, while Millikan is sceptical about the capacity of animals to represent "dead facts", with no evident practical relevance, as inconsistent, hence about

their capacity for theoretical rationality.

1.2.5. Hurley. Philosopher Susan Hurley pursues the possibility of domain-specific practical rationality. She argues that practical rationality does not require conceptual capacities. In one common understanding, conceptual capacities require the ability to decompose and recombine the elements of a thought in inferential patterns with unrestricted generality, and hence underwrite a domain-general reasoning capacity, or "inferential promiscuity". It may be implausible to attribute conceptual capacities in this sense to non-linguistic animals41, but that does not mean that they cannot be correctly understood as acting for reasons. It is possible for animals to act for reasons even though their reasons for acting are bound to specific domains, so that only certain aspects or subsets of their behaviour is rational. Animals can occupy "islands of practical rationality".

Hurley follows Daniel Dennett and others in distinguishing processes at the sub-personal level – where we describe or explain an action in terms of causal mechanisms, in neural, structural, or functional terms – from processes at the personal level – where we make sense of actions in terms of holistically related and normatively constrained mental states. At the personal level, we understand intentional agency in terms of the rational interactions between someone's various desires and her beliefs about how best to satisfy them. Hurley agrees with John McDowell (1994a,b) that personal-level explanations of behaviour in terms of reasons need not correspond to or compete with subpersonal explanations in terms of causal mechanisms, even though the latter enable people to behave in ways that make personal-level descriptions appropriate. She also holds that our grounds for applying personal-level descriptions are not the structure of internal subpersonal mechanisms, but patterns of observable behaviour in environments; minds are extended, and rational processes are widely distributed. This is realism about minds and reasons, but of a strongly externalist stripe. Hurley suggests an animal analogue of the personal/subpersonal distinction, and argues that animal action can sometimes be properly understood at the animal rather than the sub-animal level. The weight of her claim that animals can act for reasons rests on the view animal action can sometimes be properly understood as displaying holism and normativity, despite their lack of fully-fledged conceptual abilities.

To defend this view, Hurley first explains how holism and normativity could obtain despite a lack of domain-general conceptual abilities, and then illustrates this possibility with several examples involving chimpanzees.

Hurley emphasizes that a capacity for flexible generalization is not all or nothing. The holism of intentional agency is an instrumental form of the capacity for flexible generalization that does not require conceptually structured inferential promiscuity. An intentional agent's means and ends can decouple and recombine; a given belief or a given means can serve various ends, and a given desire can be served by various beliefs and various means, depending on circumstance. An intentional action results from the interaction of the agent's ends and his perception of means/end contingencies, so that an effective means to his ends in given circumstances is selected from those available. But flexibility in adopting effective means to ends can be a matter of degree; some skills that are used in instrumentally rational ways in one domain may not generalize beyond that domain. For

example, an animal may be able to act in accord with transitive inferences about social dominance in selecting effective means to its ends, but be unable to act in accord with transitive inferences about foraging opportunities. It may not have the conceptual abilities needed to make this transition.

The normativity of acting for reasons requires the possibility of mistake: that there be a difference between making a mistake in acting for one reason, and acting for a different reason. (Note that this is a weaker sense of normativity than, for example, Dretske's: one can make a mistake in acting on a bad reason.) Hurley notes that this in effect requires a solution to Kripke's (1982) version of the rule-following problem, applied to intentional agency. This difficult problem arises equally for subjects with conceptual, theoretical, and linguistic skills, and Hurley doubts it is more tractable for them than for mere intentional agents: it isn't clear why the finer-grain, domain-free flexible generality of inferential promiscuity secures any advantage here. Rather, Hurley suggests that normativity depends on complex control and simulation capacities embedded in teleological contexts, which leave the possibility of mistake no less problematic for mere intentional agents than for creatures with conceptual abilities.

Hurley illustrates the possibility of context-specific reasons for action by reference to experimental work with chimpanzees. The first example illustrates how a symbolic context might make instrumental reasons available to an animal, which it cannot generalize to similar but non-symbolic contexts. (The example is drawn from experimental work by Sarah Boysen described in her chapter in this volume.) Sheba is faced with two dishes of candy, one of which contains more candies than the other. When she points to one dish, she reliably is given the candies in the other dish. Despite this, she continually points to the dish with more candies, so is given the smaller number of candies. In a variant of the task, different numerals are placed in the dishes instead of candies; Sheba has already learned to use numerals in other tasks. When she points to a dish with one numeral in it, she is given the number of candies signified by the numeral in the other dish. She at once begins pointing to the dish with the smaller numeral, and is thus given the larger number of candies. The symbolic context appears to makes instrumental reasons available to the animal. Early studies suggested that this instrumentally rational behaviour would persist when one dish contained a numeral and the other candies (compare Boysen's chapter in this volume for later work). However, when both numerals are replaced with candy again, she reverts to pointing to the dish with more candies. Sheba appears to point to the dish containing the smaller numeral in an instrumentally rational way, order to get as many candies as possible, but cannot generalize this instrumental reason to the all-candy case. Hurley discusses the interpretation of this task in comparison to a related task given to children.

The second example illustrates how certain social contexts might make reasons for action available to an agent that it cannot generalize to other contexts in which they are also relevant. Various paradigms have been developed to test for mind reading in non-linguistic animals (these methods are discussed further in the chapters in Part 5 of this volume). In the hider-communicator task, success requires attribution of a false belief to a helpful communicator who indicates the location of food, in circumstances where the communicator has been 'tricked' by a third party. This test produces results in children that correlate well

with the results of standard verbal false belief tests of mind reading. Chimps fail this test of mind reading. In a different paradigm, attributions by a subordinate chimp to a dominant chimp of perceptions concerning the presence of food would enable the subordinate to compete more successfully in acquiring food. Chimps pass this test of mind reading; subordinate chimps appear to use information about what dominants did or did not see in instrumentally rational ways. Why should chimps be able to use information about the mental states of others in instrumentally rational ways in one paradigm but not the other? One possibility is that their ability to treat information about the mental states of others as reasons for action is specific to social contexts involving competition for food, which are ecologically significant, but do not generalize to social contexts involving cooperation over food, which are not ecologically significant.42

Hurley concludes by explaining how her view of domain-specific practical reasons shares ground with simulationist accounts of cognitive capacities, and by arguing that the functions and interest of making sense of action are not wholly discontinuous at the boundary between human and non-human animals. We should, she claims, avoid over-intellectualizing what it is to have a mind, or viewing rationality as all or nothing. By allowing that rationality can be disaggregated into domain-specific capacities within which basic features of practical rationality are nevertheless present, we can characterize the animal level in terms that are neither too rich nor too impoverished, and chart various specific continuities and discontinuities between our minds and those of other animals.

Hurley is blithely compatibilist about personal level rational processes and subpersonal processes that don't map tidily onto the former.43 Such compatibilism can be challenged. Explanatory subpersonal mechanisms, such as associative mechanisms, are often regarded as competing with and threatening to displace rational explanations. If so, then even if rational explanations of human behaviour can be defended as indispensable for many purposes, the same may not be true for non-human behaviour. This issue is discussed in more detail in the contributions to part II.

1.3. Rational versus associative processes.

A powerful form of subpersonal explanation of behaviour in both human beings and other animals appeals to associative processes, including classical and instrumental associative conditioning processes and the associative neural networks that may support them. The authors in this section are all concerned with the contrast between explanations of behaviour in terms of associative processes and explanations in terms of "higher" processes: cognitive, or rational, or intentional explanations. Do such contrasts hold up, and if so, do associative explanations compete with and displace the higher forms of explanation?

1.3.1. Allen. Philosopher Colin Allen focuses on transitive inferences, of the form: A > B; B > C; therefore, A > C. Allen compares cognitive and associative accounts of behaviour by animals on transitive inference tasks. In doing so illustrates a more general methodological dispute between ethological and experimental approaches.

Transitive relationships are often important to animals, especially social animals. Transitive

inference would permit, for example, an animal to behave in a way appropriate to the dominance relation between other animals A and C, even if their relationship has not been directly ascertained; this could enable the animal to avoid injury and loss of various resources. Given the ecological significance of transitive inference in certain domains, we might expect to find domain-specific rather than domain-general capacities to behave in accord with such inferences. But even if such behavioural capacities were limited to a specific domain, the question would still arise whether they were best explained cognitively or associatively. Cognitive explanations postulate that animals explicitly represent orderings such as A>B>C>D>…, and use these representations to infer relationships between non-neighbouring pairs. By contrast, associative explanations appeal to past reinforcement histories of the individual elements, without invoking any explicit representation of the entire series.

Allen describes the ping-pong game between cognitive and associative accounts of apparent transitive inferences. If A is rewarded rather than B and B is rewarded rather than C, pigeons and rats given a novel choice between A and C are likely to choose A. This is easy to explain associatively, since A has always been rewarded and C never. An extended version of the test trains animals on five instead of three elements: A is rewarded rather than B, B rather than C, D rather than D, and D rather than E. Many animals given a novel choice between B and D reliably choose B. The original associative explanation no longer applies, since D has been rewarded in training just as often as B. However, more sophisticated associative explanations appeal to value transfer effects: the sure-thing value of A rubs off by association onto B. There is experimental support for this account, though ambiguity between cognitive and associative accounts of this and other experimental results remains.

Animals in experimental paradigms that test for transitive inference learn the relevant associations slowly and laboriously. Ethologists may wonder how relevant the results of such experiments are to explaining the capacities of animals living in complex natural societies, where dominance hierarchies including many dozens of animals are rapidly learned and changes in dominance are flexibly adjusted to. Allen suggests the slow learning rate in the experimental transitivity tasks may be due to the specific nature of the tasks: they use arbitrary stimuli of no ecological significance – such as shapes shaded with patterns – and the ordering of them doesn't reflects a natural or biologically significant ordering, but simply the consequences of a training paradigm. By contrast, the social dominance relations that monkeys negotiate with ease have clear biological significance. Could capacities for transitive inference be domain-specific?

If so, what might that suggest about the general distinction between cognitive and associative explanations? Are these exclusive alternatives? Are associative explanations, when available, always to be preferred? Is scepticism warranted about the defensibility or utility of the general distinction (as expressed by Papineau and Heyes in their chapter)? Allen suggests that discussion of this last question may be too abstract to be instructive. Cognitive processes may bear family resemblances to one another rather than satisfy an analytical definition. And even if the general cognitive/associative distinction cannot be defended, a distinction between different kinds of explanation of specific capacities, such as the contrast between cognitive and associative accounts of behaviour in accord with transitive inference, is still tenable and useful. An explanation that postulates a representation of an entire ordered

series and associated cognitive processes that exploit such a representation is a quite different explanation from one in terms of associations between stimuli and responses of varying strengths. We should keep in mind, however, that cognitive accounts of transitive inference may be correct only in application to specific domains where natural orderings and transitive inference are biologically significant, while associative mechanisms play a role in explaining other behaviours that conform to transitive inference. Moreover, cognitive and associative mechanisms for such behaviours may co-exist in one animal, each applying to different domains, and yielding different behavioural capacities in experimental tasks that lack ecological significance as opposed natural social environments.

1.3.2. Papineau and Heyes. Strong scepticism about the heuristic and theoretical importance of a distinction between rational and associative explanations is expressed by philosopher David Papineau and psychologist Cecilia Heyes (chapter 8). Indeed, they suggest that the rational/associative dichotomy is a modern descendant of Descartes' soul/matter dichotomy, and equally unhelpful; it obscures the real issues. There is no Rubicon between associative and rational processes to be crossed; rather, evolution adds specific new cognitive capacities by tinkering with previous mechanisms. What we want to know is how animals work: which specific mechanisms account for which specific cognitive capacities in which animals.

They illustrate their view by reference to experiments on imitation in Japanese quail (by Thomas Zentall and others), describing another ping-pong game between rational and associative explanations. The behaviour to be explained is as follows. Observer quails watched a demonstrator quail push a lever; some demonstrators pushed by pecking and others by stepping. Half of the observers saw a demonstrator rewarded with food for this behaviour, while the other half saw a demonstrator perform the same behaviours without reward. Only those observers who had seen the demonstrator rewarded imitated its behaviour when later given access to the lever themselves: they pecked if their rewarded demonstrator had pecked and stepped if it had stepped. The observers of the unrewarded demonstrators did not differentially peck or step. The observers were thus sensitive to demonstrator reward. They appear to learn imitatively, in the sense of learning how to achieve an end by observing another's actions.

It is tempting to explain this sensitivity in rational terms. That is, the imitator birds want food and believe that doing the same thing as the rewarded demonstrators will provide it:

X's pecking produces food

My pecking will produce food

I want food

Therefore,

I will peck

And it can also be hard to see how this behaviour could be explained in terms of standard

associative learning. To see why, note that there are four elements present: demonstrator behaviour, demonstrator reward, observer behaviour, and observer reward. First, how is demonstrator behaviour associated with similar observer behaviour? What the quail sees and feels when it pecks is very different from what it sees and feels when the other pecks. Perhaps this correspondence problem can be avoided by some associative mechanism or innate tendency for mimicry. But second, that would still not explain why the observer mimics only rewarded behaviour. Neither instrumental nor classical conditioning readily explains why an association between demonstrator reward and demonstrator behaviour, plus an association between demonstrator behaviour and observer behaviour (expressed in mimicry) would yield an association between observer behaviour and observer reward. There is nothing obviously rewarding to the observer in seeing some other bird getting food. For related reasons, a capacity for imitative learning, to absorb new means-ends information by observing others, is widely regarded by scientists as a hallmark of advanced cognitive capacities (despite the erroneous popular view of imitation as a low-level capacity).44

The apparent difficulties of explaining the quails' behaviour in associative terms may seem to reinforce the rational explanation. This would not entail unlimited cognitive powers; the quails' ability to act rationally on the basis of their beliefs and desires may be specific to certain domains. But even if domain-specific rational explanation is admitted, the issue is not yet resolved. Associative learning theory has further resources for explaining sensitivity to demonstrator reward. Quails tend to feed together, so a quail's feeding may be associated with observing others feed with the result that seeing another feed acquires rewarding properties. Add to this the tendency to mimicry, so that seeing another peck is associated with pecking oneself. When the observer sees the demonstrator peck and feed, that secondary reward is associated with its own pecking in mimicry, differentially reinforcing mimicry of rewarded demonstrations.

Can an experiment decide between this improved associative account and the rational account? Papineau and Heyes describe a hypothetical variant of the quail experiment, using some foods that are devalued (by pre-feeding the birds to satiety) and other foods that are not, which might seem able to do so. But they go on to explain how "rational" behaviour in their experiment could also be explained in terms of further associative mechanisms, using a cybernetic model of instrumental learning that associates a response with its outcome, including devalued foods, in a negative feedback loop.

Rational explanations of behaviour are typically distinguished from associative explanations by featuring norm-governed reasoning involving belief-like representations. But Papineau and Heyes doubt that the rational/associative distinction can be clearly drawn by reference to the processing of representations. If rational processes are defined in terms of representations, there is no good reason not to include associative processes as rational. Associative learning involves modifications to internal states that carry information, and the informational processes involved in sophisticated forms of instrumental conditioning are analogous to those in practical inference such as: doing B in C will lead to D; I am in C; I want D, so I do B. Admittedly, associative informational processes may be limited in the kinds of inferences they can enable. But if rational processes are required to underwrite unlimited inferential powers, they represent a cognitive ideal that is useless as a research tool

applied to animals—whether human or nonhuman--, whose cognitive capacities must be realized by specific mechanisms.

The message from Papineau and Heyes is: we should refocus on specific explanations of how animals do specific things, rather than on the presence or absence of some general or ideal form of rationality that contrasts with associative mechanisms. Perhaps language makes ideal rationality accessible to human beings – though proponents of bounded and ecological rationality would dispute this also. The recommended refocus may not just change our approach to animal minds, but also to assessing the continuities and discontinuities between animal minds and our own.

1.3.3. Clayton, Dickinson, and Emery. Papineau and Heyes are careful to say that there are no obvious associationist explanations of some of the impressive cognitive feats of scrub jays described by psychologists Nicky Clayton, Tony Dickinson, and Nathan Emery (chapter 9). Clayton and colleagues contrast mechanistic explanations of behaviour in terms of associative processes with intentional explanations in terms of rationally flexible interactions between beliefs and desires (as in Kacelnik's PP-rationality, or Dretske's minimal rationality). Explanations in terms of beliefs and desires are intentional in the sense that they depend on the representational content of beliefs and desires, their truth or fulfilment conditions. By contrast, associative explanations depend on mechanistic properties. Associative learning consists in the formation of excitatory or inhibitory connections between nodes activated by various events; excitation is transmitted from one node to another until activation of the terminal node is sufficient to generate the observed behaviour. Clayton et al view these different styles of explanation of a given behaviour as mutually exclusive, although different behaviours of the same animal can be explained in different ways. Rationality is primarily a property of processes causing a particular behaviour; a rational animal, whether human or non-human, is an animal some of whose behaviour—not necessarily all—warrants intentional explanation.

Associative processes may seem capable of explaining memory for food caches in scrub jays. Nodes activated by visual cues around the cache site become associated with those excited by food stored at the site. Re-exposure to the cache site's stimuli activates food nodes that are associated in turn with nodes controlling the appropriate response. But Clayton et al argue that the food-caching and retrieval behaviour of the birds they study is best explained intentionally. Given a certain desire for food and a belief about where the food is, the rational thing to do is to search in certain cache sites. The rationality of such practical inferences, which depend on the contents of desires and beliefs, distinguish intentional explanations from associative explanations, even if the latter are adaptive (and hence biologically rational, in Kacelnik's sense). While intentional accounts are often intuitively attractive, Clayton et al emphasize that behaviour does not wear its intentionality on its sleeve; rival associative accounts are often available. However, rational explanations of the behaviour of non-linguistic animals are more empirically testable, with the right experimental procedures, than sceptics may suppose.45

(a) Philosophers have sometimes argued on very general grounds that intentional states cannot have determinate content in the absence of language—or at any rate that we could not

know what their contents are. The first line of experiments described by Clayton et al shows how it is possible in specific cases to determine the content of desires on the basis of non-linguistic behaviour. Their empirical procedure addresses whether a desire has a specific or a general content: is it a desire for food, say, or a desire for peanuts? Having cached two different types of food the birds are, after an interval, fed to satiety on one type and then allowed to search their cache sites. They selectively search sites where they cached the other type of food. This behaviour is better explained in terms of a desire for a specific food rather than general desire for food. While an associative account of this behaviour can be offered, in an experiment with a more complex design the birds behave as predicted by an intentional rather than an associative account. There may never be a once-and-for all, decisive test of the intentionality of animal behaviour, but if we focus on testing particular predictions of associative versus intentional accounts in specific cases the issue is not empirically intractable.

(b) The second line of experiments described concern whether jays' caching behaviour should be explained in terms of the flexible interactions of beliefs based on various declarative memories in practical inferences. Human cognitive psychology contrasts a declarative memory system, in which general knowledge and recollections of specific episodes interact rationally and flexibility in guiding action, with procedural memories embodied in skills, responses, and habits that are often domain-specific. Declarative memories in jays might include both general memories, say concerning the location of reliable food sources, and specific episodic-like memories of particular life events, such as what food they cached where and when. The scrub jays studied by Clayton and colleagues eat a variety of foods, but have favourites; they prefer crickets to peanuts, for example, but only when the crickets are fresh; and crickets decay relatively rapidly, while peanuts do not. The jays were tested to see if their behaviour in searching for cached food can be explained by flexible, rational interaction of general declarative knowledge of how rapidly crickets decay with specific episodic-like memories of when and where they cached crickets as opposed to peanuts.

The tests proceeded as follows. Birds were allowed to retrieve food after one day and then after four days (no other intervals were used in the initial learning trials); they rapidly learned to search for cricket caches after one day, when the crickets were still fresh, and for peanut caches after four days, by which time the crickets were decayed. Next the experimenters pilfered all the cached food, and then allowed the birds to search after one, two, three, four, or five days. The jays searched for cricket caches at intervals up to three days and after longer intervals switched to searching for peanuts. This behaviour suggests that they had generalized from their experience of crickets fresh after one day but not four days, and formed a general belief that crickets stay fresh for only three days, which interacted in instrumentally rational ways with their memories of specific caching episodes to guide their searches.

But how flexible was this interaction? A process of practical inference operating on the contents of declarative memories would produce an instrumentally rational change in behaviour if the content of general declarative memories changed in certain ways after caching had occurred.47 To address the issue of rational flexibility, Clayton et al allowed the birds to find cached food for the first time after three days and to learn, contrary to their prior

generalized belief, that crickets had in fact perished by this time. A revised general belief about decay rates should rationally lead to a reversal of behaviour: namely, to searching for peanuts instead of crickets after three days. This is indeed what the jays did. This behaviour reversal can be explained as instrumentally rational, in terms of the flexible integration of the contents of their food preferences and their recollections of specific caching episodes with the contents of their changing general beliefs about decay rates of specific foods.

(c) Recall that the demands of social interactions, with the potential to produce manipulation of information, have often been viewed as driving the evolution of advanced cognitive capacities. The third line of experiments investigates caching behaviour in social contexts. Many birds are known to pilfer the caches of other birds they have observed caching, and corvids have been observed to return alone and re-cache to new sites if they had been observed by other birds during the initial caching. But such behaviour in the wild may be coincidental, and does not resolve whether re-caching should be explained intentionally. In an experiment, jays re-cached significantly more items in new sites when they had been observed during caching than when they had cached in private. Can this behaviour be explained mechanistically? Perhaps re-caching is automatically triggered by the memory of having recently been in the presence of another bird. This mechanistic explanation is ruled out by another experiment. Birds who cache in an "observed" tray immediately before or after caching in a "private" tray, and are then allowed to re-cache from either tray, re-cache selectively from the observed tray rather than the private tray.

An obvious intentional explanation of the selective re-caching is that that birds have general beliefs that being observed during caching tends to lead to loss of food and that such loss can be prevented by re-caching, which interact with their memories of specific caching episodes and their desires to avoid loss of food. But how might such general beliefs arise? A final study compares re-caching by birds who are experienced thieves themselves with re-caching by naïve birds who have watched other birds cache but have no experience of pilfering other birds' caches. The experienced thieves selectively re-cached to new sites when their initial caching had been observed; the naïve birds did not. A rational explanation of this behavioural difference involves a kind of mind reading: the experienced thieves appear to attribute their own belief, that observed caches can be stolen, to other birds who observe caching. On this account, their selective re-caching of observed caches to new sites is explained by the rational integration of their own experience of pilfering with memories of specific episodes in which they cached food while observed.

The converging evidence for intentional rational processes from these three lines of experiment is impressive indeed. This work contributes strikingly to the growing body of evidence that "bird brains" may be small but have been seriously underrated: under rigorous experimental probing, birds can compete with the best in the animal rationality stakes.

1.4. Metacognition

One approach to distinguishing rational processes from other cognitive processes focuses on metacognition. If rationality requires that the agent can recognize the possibility of mistake, then metacognition could satisfy this demand by allowing the subject to monitor its own

cognitive processes. Metacognitive processes would give agents information about the information on which they act, enabling them to monitor certain behaviour-producing processes to check whether, in the circumstances, they are likely to lead to success (as in Millikan's trial and error in thought; see chapter 4). Domain-specific forms of metacognition may evolve as adaptations to particular environmental problems, and rational processes may eventually emerge from cumulative metacognitive adaptations.48 The contrast with associative processes arises again here: when is it right to explain animal behaviour in terms of metacognitive as opposed to associative processes? Can metacognition itself be explained as emerging from associative processes, and hence provide an evolutionary bridge between associative and rational processes? What kinds of non-linguistic behaviour would provide evidence of metacognition? An evolutionary perspective is helpful in understanding how associative, intentional, and metacognitive processes may co-exist in a given animal, including in human beings: an animal whose behaviour is best explained intentionally or metacognitively in a given domain may fall back on associative processes in other domains (see Call's chapter)--conceivably these different processes might even compete to govern a given type of behaviour in some circumstances. What implications might such an evolutionary perspective on metacognition have for human rationality? Is the final linguistic adaptation by the human animal just another, if rather special, metacognitive layer, or does it make a qualitative difference in enabling classical domain-free reasoning capacities?

1.4.1. Call. Psychologist/primatologist Josep Call (chapter 10) presents a variety of behavioural evidence that he argues is best explained either by causal reasoning or by metacognition, and goes on to consider the relationships between capacities for reasoning and for metacognition.

(a) Causal reasoning. Call first argues for a view exactly contrary to the widespread view that animals are excellent at making associations between arbitrarily related stimuli, but poor at reasoning and causal inference.49 Perhaps ironically, the greater ease with which human beings learn arbitrary associations may be linked to human facility with symbols, which are arbitrarily associated with what they represent. Call describes performances by various great apes in a series of experiments that contrast arbitrary associations with causal associations, using tasks with similar superficial cues and reward contingencies but different causal structures. The apes do significantly better on the causal tasks than on the arbitrary tasks.

For example, in a causal task the apes must select from two cups the one that is baited in order to be rewarded; both cups are shaken, and only the baited one makes noise—the noise is caused by the food inside the cup. In the paired arbitrary task, both cups are tapped but only the baited one makes a loud tapping noise—a noise causally unrelated to the presence of food in the cup. Apes select the baited cup above chance in the causal task but not the arbitrary task, suggesting that they infer that food in a shaken cup would make noise and do not simply associate food with noise. In a revealing variant on the causal task, the empty cup was shaken and the baited cup was silently lifted; neither produced noise. A negative causal inference is suggested: if food causes noise in a shaken cup, then the silence of a shaken cup implies that food is in the other cup. Apes acted in accord with this inference: they tended to select the lifted cup in this task (when compared with a control task in which both cups were lifted). Similar results supporting causal inferences rather than associative processes were

found in visual causal and arbitrary tasks. In the causal task apes were rewarded for selecting an inclined board propped up by food beneath it instead of a flat board, while in the arbitrary task they were rewarded for selecting a previously examined wedge (with the same angle of inclination and presenting the same visual aspect as the inclined board in the causal test) instead of a flat board. Apes selected the inclined board hiding food in 80% of the trials but the wedge in only 50% of the trials, suggesting that they understood that the inclination in the causal task but not the arbitrary task was caused by the presence of food. Strikingly, there is no evidence of learning in these tests: the apes either performed well from the beginning (in the causal tasks) or poorly throughout (in the arbitrary tasks).

These results suggest that apes do not simply associate a cue with the presence of food, but understand that the food is the cause of the cue, and can reason accordingly. They could be regarded as providing empirical evidence of a capacity for proto-logical reasoning similar to that described by Bermúdez, relying on pairs of contraries (for example, noise versus silence) and conditional dependence based on causal dependence (for example, if a cup containing food is shaken, noise results). Call argues that rival explanations of these results, either in terms of associative learning and reinforcement, or in terms of innate dispositions, are unpersuasive. However, these different processes may co-exist.

(b) Metacognition. Do animals know whether or not they know how to solve problems? In a second line of argument, Call presents behavioural evidence for metacognitive processes in certain animals—but not others, describing recent results from two experimental paradigms.

The first paradigm allows animals to escape from tasks if they are uncertain of the correct response, where escaping produces less desirable outcomes than responding correctly. For example, an auditory task rewarded discrimination of high from low pitches by pressing "high" and "low" keys, respectively, but also presented an "escape" key which moved the subject on to the next trial (with increasing delays to discourage overusing). Dolphins, like human subjects, used the escape key for difficult intermediate pitches. In a visual delayed match to sample task, monkeys did better when allowed to use the escape key than when they were not given this option, suggesting that they knew when they had forgotten the sample stimulus.

The second paradigm allows animals to seek additional information before responding when they are initially given incomplete information. Subjects – including two-year-old children, orangutans, chimpanzees, gorillas, and bonobos – faced two opaque tubes, and were rewarded for touching the baited tube. In the visible condition the experimenter placed food into one tube in full view of subjects; in the hidden condition subjects knew one tube was baited but did not know which. Subjects were allowed to look inside the tubes before choosing one, and did so more often when they had not seen the baiting. Moreover, they often chose the other tube as soon as they had looked into an empty tube, without looking to see the food in the other tube. These results suggest both metacognition and inference: apes know that they do not know where the food is, so seek additional information, and when they find that the food is not in one tube they then infer that it is in the other. Dogs, by contrast, failed this test. The results from these two paradigms suggest that apes can monitor whether they have the information needed to solve a problem and either escape the situation or seek

more information accordingly.

In conclusion, Call suggests that cognitive capacities for causal reasoning and metacognition should be regarded as evolved adaptations. These cognitive adaptations need not wholly displace associative processes; the latter may represent a fallback mechanism for apes when their capacities for reasoning and reflection cannot be engaged. Call is tempted by the view that reasoning and even linguistic capacities may build on metacognitive capacities in a co-evolutionary process.

1.4.2. Shettleworth and Sutton. Can success on behavioural tests for metacognition can be explained in other, deflationary ways? What kinds of controls are needed to rule out alternative explanations? These methodological questions are probed by psychologists Sara Shettleworth and Jennifer Sutton in their review of metacognition paradigms and results with animals (chapter 11). In human beings, success on metacognitive tests correlates with subjects' report of a "feeling of knowing". Accurate metacognitive awareness has obvious functional advantages: if one can monitor information gaps, one can take steps either to avoid situations that require the missing information or take steps to acquire it. While we may never be able to assess whether animals have a feeling of knowing, with the right controls we can define functionally similar behaviour and thus assess whether animals' performance depends on metacognitive information.

Nonverbal metacognitive paradigms generally build on well-established nonverbal paradigms for testing perceptual discrimination and memory. The discrimination or memory tests vary in difficulty, and to them is added an "escape" option, which carries a reward intermediate between correct and incorrect responses on the discrimination or memory task. But, as Shettleworth and Sutton explain, success on some such tests allows of a deflationary associative explanation. For example, suppose an animal consistently escapes from memory trials that are more difficult because they involve a longer interval between stimulus and memory test, but accepts easier trials with a shorter interval. Is this because the animal knows it has forgotten in the harder trials? Or is it because it has simply learned associatively that its rewards are maximized by taking the escape option after long intervals and the test option after short intervals? Controls must be added to such a behavioural test to define functional similarity to metacognition more tightly and rule out alternative explanations.

One important control in metacognitive paradigms is randomly intermixed forced trials, where the escape option is omitted. The metacognitive account predicts that an animal will do worse on the forced trials than on trials it chose to accept when the escape option was available. Another second important control is to transfer the animal to tests with new stimuli, so that it cannot rely on learned associations between particular stimuli and rewards, to see if the results across different tests converge on the metacognitive explanation. For example, instead of varying the difficulty of a memory task by varying the retention interval, the interval of exposure to the stimulus could be varied. The metacognitive account would predict an immediate transfer of behaviour consistent with metacognition, without the need for learned associations.

As explained also by Call, discrimination/escape tasks with dolphins (auditory discrimina-

tion) and monkeys (visual discrimination) produced results consistent with metacognition. The results were similar to those obtained from human subjects in almost identical tasks, in which people reported that they escaped trials when they felt uncertain. However, Shettleworth and Sutton point out that these studies with dolphins and monkeys did not include forced trial or transfer test controls. Thus the animals may simply have learned that certain responses to certain stimuli maximized their rewards. When pigeons were given similar tests that included controls, the results were not what metacognition would predict. The pigeons' accuracy was no worse on forced than chosen trials, and they did not differentially escape difficult trials when transferred to new stimuli. Their behaviour was better explained in terms of learning the most profitable response to various stimuli than in terms of choosing to take the test when they knew that they knew the correct response: the pigeons' behaviour was biologically rational, but not rational in a more demanding sense requiring access to the information on which they act. Thus, the evidence is against metacognition for perceptual discrimination in pigeons, and as yet inconclusive in dolphins and monkeys.

Tests of metacognition for memory, or metamemory, use two different paradigms. Serial probe tasks display a series of images, followed by a single probe image; the task is to say whether it was in the series. In more difficult trials, the probe image occurred earlier in the series, or the series is longer. Monkeys took the escape option more often in difficult trials, and were more accurate on chosen than on forced trials. These results appear to be best explained by metamemory: monkeys have information about whether they recognize the presented probe image. Delayed match-to-sample metamemory tasks are even more demanding. They require that the monkey decides whether or not to escape after seeing a sample but before being presented with the stimuli from which he chooses a match to sample. That is, the monkey has to decide how well it remembers the sample, or to predict whether it will recognize it, before seeing any further stimuli that it might or might not recognize. Metamemory tests with monkeys using this paradigm included both transfer controls and forced trial controls, and the results were as predicted by a metamemory account. The results of these two sets of tests of metamemory provide strong evidence that monkeys can monitor their own memories. By contrast, negative results on related tests with pigeons indicate that pigeons do not have metamemory.

Shettleworth and Sutton emphasize that metacognition for perceptual discrimination is distinct from metacognition for memory. A given species may have one but not the other; different species may have metacognitive capacities in different domains. Moreover, the extensive training these tests of metacognition require contrasts with the spontaneous uses of metacognition by people in everyday life. Tests for metacognition in animals might do well to focus on biologically relevant contexts in which such abilities might have a natural function. In this respect, they cite approvingly Call's paradigm in which apes can choose whether to obtain more information by looking into opaque tubes for bait before choosing one. While there is strong evidence for metamemory processes as opposed to mere biologically rational behaviour in primates, a challenge for the future is to develop more naturalistic contexts in which animals might display possibly domain-specific capacities to access the information on which they act.

1.4.3. Proust. The perspectives on metacognition offered by Call and by Shettleworth and

Sutton lead to the thought that evolution may have selected for domain-specific metacognitive capacities, in accordance with their adaptiveness to different animals. A related evolutionary approach to metacognition is given a rich theoretical framework by philosopher Joëlle Proust (chapter 12). Proust contrasts top-down approaches to animal rationality, which begin with human reasoning and look for something relevantly similar in nonhuman animals, with bottom-up approaches such as her own, which view human rationality as emerging from evolved mechanisms for control and representation in non-human animals. Her synthesizing discussion charts the adaptive steps that lead from merely biologically rational behaviour to human reasoning processes, and clarifies the critical functions of metacognitive processes in this genesis.

(a) Rationality as cognitively-operated adaptive control. Proust finds a powerful clue to the biological sources of rationality in the very idea of bounded rationality: the limited informational resources of simple cognitive systems in variable environments bring with them new selection pressures for flexible behaviour. These favour capacities to extract information from changing environments so as to respond appropriately, and to control and manipulate informational resources in ways that advantage oneself and disadvantage one's competitors. She endorses the view of Peter Godfrey-Smith (1998, 2002) and Kim Sterelny (2003, and Sterelny's chapter herein) that variable, complex environments exert strong selective pressure for flexible behaviour.

Adaptive control structures evolve in response to this pressure. Information flows in a dynamic loop through control structures: given a target, the system generates output, which in turn generates feedback, which is compared with the target; the comparison of target and feedback governs output in a way that permits disturbances to be neutralized. In "slave" control systems such as thermostats,50 this comparator process is implemented inflexibly by the physics of the system; they cannot learn or change their goals. By contrast, adaptive control systems achieve flexibility by the use of internal models: learning and memory allow them to predict the results of alternative commands and flexibly to select the right commands to achieve their target in varying environments. Simple internal models can be local, specialized, and not refer to anything outside the system. But control becomes more powerful and flexible when internal models employ cognitive representations, which are selected to refer to specific events or properties that can be re-identified over time and which can combine in indefinitely new ways to represent alternative possible courses of action.

In its earliest form, Proust proposes, rationality can be understood as a disposition that tends to be realized by adaptive control systems that are cognitively operated. Her notion of cognitive operated adaptive control is closely related to Dretske's conception of minimal rationality, in which the explanation of behaviour depends on the content of representations, as well as to Clayton et al's conception of intentional explanation. However, such flexible control has costs as well as benefits, which vary with both the ease with which information can be extracted and its reliability; in some environments, rigid behaviour can bring greater rewards than flexible behaviour, which may involve, for example, higher chances of error. Moreover, in the variable, informationally noisy, environments that select for flexible cognitive control, information itself becomes a good. Organisms face various trade-offs and constraints in discriminating signals from noise and in categorizing signals. They are

advantaged if they can manipulate the informational properties of their environments effectively, to increase information quality for themselves and decrease it for their competitors.

(b) Rationality as metacognitive control. What Proust calls "epistemic actions" are actions with the goal of such informational manipulation. This goal can be pursued by physical means, such as squirting ink in the eyes of a predator, or by representational means, such as issuing a deceptive signal. Proust refers to epistemic actions pursued by representational means as "doxastic actions", and identifies metacognition as a sub-category of doxastic action of particular relevance to the development of rationality: metacognition pursues informational goals in the system itself by informational means. It monitors and controls various of the system's own cognitive processes (as opposed to controlling behaviour directly), functioning to improve cognitive flexibility and informational quality in complex, multi-valued control systems. However, metacognitive control processes also bring a new set of costs, such as the need for complementary mechanisms of binding, calibration, inhibition, and so on. Whether metacognition is worth the cost in the sense of biological rationality varies with the needs and environment – including informational environment – of different species.

Metacognition is covert in the sense that it enables a creature to make trials and errors in its head, as Millikan puts it – to entertain possible actions without incurring the potential costs of actual actions. Metacognitive capacities emerged, Proust suggests, in the form of covert simulative processes, which use the action control system itself off-line to represent possibilities (the need for inhibitive mechanisms in off-line simulation is one of its costs). Like Hurley, Proust views simulation in control systems as an efficient and adaptive way of implementing a variety of cognitive functions; there is increasing evidence that nature has used it widely. Note that while simulative metacognition can improve the results of my actions here and now, it also enables various forms of decentering from me-here-and-now, which we (along with various contributors) have suggested is an important aspect of rationality.

(c) Rationality as explicit metarepresentation of reasons. Proust next highlights an important distinction between metacognition as a practical capacity for information control exemplified by simulation, and metarepresentation, a theoretical capacity for explicit, conceptual representation of mental states such as beliefs and desires—one's own and others (see also and cf. Bermudez 2003, ch. 9, 166ff). A closely related distinction has played an important role in debates between theory-theory and simulation-theory concerning the basis of mind reading capacities. She surveys empirical work with animals (including work discussed by other contributors) which provides evidence that metarepresentation is not required for metacognitive control, of either information available to others or to oneself. On the other hand, there is no species with capacities for metarepresentation but not for metacognitive control. She explains this asymmetry in terms of a distinction between the implicit, procedural reflexivity of metacognitive control and the explicit, conceptual reflexivity of metarepresentation. The continuous dynamic cycles of control architectures, including architectures for metacognitive control, mean that command and feedback refer to one another procedurally or implicitly: metacognitive information is embodied in the dynamic

properties of the control loop used in simulative mode. This can, but need not, be articulated as explicit mental content.

How and why is implicit metacognitive information available to some animals transformed into explicit metarepresentation, which is characteristic of the final stage of fully-fledged human rationality? She suggests that in human beings the output of metacognitive control loops, which carry information about information, may be inputs to higher-level structures that control linguistic communications. The latter recode metacognitive information in an explicit, metarepresentational format. This further capacity serves social control functions in a Machiavellian world in which information is manipulated by linguistic (among other) means, by permitting one to report one's reasons and offer justifications for acting. She cites neurophysiological evidence for such a superposition of control structures in human beings; both forms of control are aspects of human rationality, but only the lower level of rationality is shared with other animals.

As we've seen, as well as metacognition, normativity is often regarded as a requirement of rationality. Proust ends her discussion of metacognition and rationality by arguing that normativity depends on explicit metarepresentation. Metacognition without metarepresentation is not prescriptively normative: it does not determine an optimal solution that the system should select. Thus non-linguistic animals cannot meet this requirement. Here she takes issue with Millikan's account of normativity in terms of natural evolutionary functions,51 including cognitive functions, according to which the functions of representations underwrite the normativity of their content. Proust denies that biological functions are natural, causally efficacious properties. Rather they supervene on the history of causally efficacious physical properties, and are relative to the explanatory concerns of biologists. Metacognition without metarepresentation can indeed support the capacity of control systems to recognize and correct mistakes relative to their purposes; but there is nothing "necessarily and inherently bad" about such mistakes. This is insufficient for prescriptive normativity in Proust's strong sense. She concludes that prescriptive normativity is only possible when metarepresentation, with its social control functions, enables explicit, public justification and assessment of mistakes.

We might raise the question, however, of why the social control functions of reporting reasons and offering justifications are any more naturally prescriptive, or any less relative to the human concerns or the purposes of control systems, than are biological control functions. Does the possibility of explicitly and publicly assessing and recognizing a mistake make the mistake necessarily and inherently bad? It is useful to distinguish social normativity from biological normativity, but it is not clear why social functions and associated norms are more objectively, inherently or naturally prescriptive – hence more normative – than biological functions and associated norms. The difference may be one of kinds of normativity rather than degree of normativity. This suggestion goes naturally with scepticism about the very idea of intrinsic prescriptivity, a scepticism that might undermine the capacity of this idea to support a conception of normativity distinctive to human rationality. But such scepticism is quite compatible with a characterization of the distinctive normativity of human rationality in terms of the social functions of explicit metarepresentation.

1.4.4. Currie. We suggested above that we are inclined to recognize rationality as opposed to "mere" intelligence when the flexibility of behaviour indicates a capacity to decentre from me-here-now. Philosopher Gregory Currie (chapter 13) suggests that since pretence depends on a capacity to decentre, it is an indication of rationality. In pretence, a creature responds to the world as transformed by imagination.

While decentering may be a form of metacognition in Proust's sense, Currie argues that it does not depend on metarepresentation. He distinguishes the vertical ascent involved in metarepresentation (as in the movement from thinking "p" to thinking "She thinks that p") from the horizontal shifts involved in decentering (as in shifts of perspective from me-here-now to me-there-yesterday or me-there-tomorrow or you-here-now or some fictional character-place-time). The vertical and horizontal dimensions of variation are conceptually indepen-dent. Representing the world from another perspective by decentering needn't necessarily depend on represen-ting another perspective – metarepresentation. Conversely, metarepresentation, say of another's thoughts, needn't necessarily involve decentering. Decentering and metarepresentation should be distinguished so that questions about whether they tend to go together and how they interact can be resolved by further empirical work and arguments. Currie doesn't try to sort out these questions in his chapter, but he does urge that issues about decentering as distinct from metarepresentation are also important for characterizing animal cognition.

What forms of behaviour provide evidence of pretence? Play need not involve the imaginative transformation and decentering of pretence. Assessing pretence often requires tracking behaviour over time and looking for elaborations of pretence that distinguish it from mere play. One such form of elaboration is recognition: for example, evidence for pretence that globs of mud are pies might be recognition that there are two more globs of mud over there, hence two more pies to be shared out. Another form of elaboration is creative enrichment, in which the game is taken to a new stage not demanded by anything so far specified or enacted. Currie takes such recognition and enrichment to be good evidence, though not strictly required, for pretence. They are found in two-year-old children, and in solitary as well as social play. But are they found in other primates?

To address this question, Currie distinguishes pretence from the overlapping categories of deception and imitation, with which pretence may be confused. Deception, in his view, requires intentional behaviour by the deceiver, directed toward an end which it is likely to achieve only by inducing a false belief, not shared by the deceiver, in the deceived. It thus doesn't require the intentional manipulation of belief, but rather behaviour that depends for its success on affecting another's belief. Deception requires first-order intentionality in deceiver (intentions) and deceived (beliefs), but not metarepresentation (intending to bring about a false belief) or mind reading. While imitation of goal-directed acts requires decentering to the other's perspective, Currie is concerned here with imitation of bodily movements, which doesn't require even first-order intentionality. Understood in these ways, imitation, deception and pretence can occur independently or together.

Various examples of animal behaviour that may seem to involve pretence are better seen as imitation or deception. A gorilla wiping a surface with no evidence of recognition or

enrichment is better seen as imitating human cleaning behaviour than pretending to clean. False alarm calls by monkeys that induce desired behaviour in conspecifics are better seen as deception than as pretending there is a threat nearby. What might warrant viewing a false alarm episode as pretence would be recognition or enrichment behaviours—say if the alarm-giver made eye movements apparently tracking the movements of a threatening predator. Eye-covering play in various primates is puzzling, but Currie sees no reason to view it as involving the imagination transformation of the world required for pretence. The few examples of recognition and enrichment behaviour in animals that do seem to warrant attributions of pretence are found human-reared primates who have been given language training and encouraged to imitate. For example, Kanzi the bonobo appeared to eat an invisible fruit and spit out its pips, signing that it was "bad", and also to grab an imaginary piece of food he had put on the floor when someone else reached out to that spot.

In Currie's view, current evidence suggests that pretence, like language but unlike deception, is beyond the capacities of non-human primates unless they are raised in such humanly enriched environments. He endorses a version of Morgan's Canon, according to which we should attribute lower mental processes that enable only observed behaviours rather than higher processes that would also enable further behaviours—unless we have some evidence of capacity for those further behaviours. Any creature that can respond to the world as imagined can also respond to the world as it is, but the converse does not hold. Since a higher mental process such as pretence enables a creature to do more, to attribute pretence we need evidence of those additional behavioural capacities, such as capacities for recognition and enrichment.

Along with Millikan, Currie allows for the possibility of rationality in perception, and shows how the kind of decentering found in pretence illustrates this possibility. Pretence, he explains, is closely linked with the perce-ptual phenomenon of "seeing-in". When you see something in a picture, you don't confuse the picture and what you see in it. Some apes, in particular human-reared apes, show evidence of seeing in without any such confu-sion and can, for example, sort line drawings according to what they depict. Similarly, we can see in simple mi-metic acts the depiction of driving a car, or nursing a baby, with no confusion of the mimetic acts and the acts they depict: we don't take them for the real thing. Currie suggests that such seeing-in is an important aspect of pretence, perhaps part of its primitive basis. If so, pretence is a kind of decentering that does not require full-blown conceptual capacities and is perceptually based: a kind of rationality in perception that is available to young children and perhaps to some human-reared non-human primates. There is evidence that what young children think is depicted or pretended depends on what they themselves see in depictions or acts of pretence.

Currie concludes with how the capacity for seeing-in and hence for pretence might have evolved. Behaviour-manipulating signals evolve into belief-manipulating deceptive signals that potentially influence a whole range of behaviours. These are in turn countered by the development of the capacity to recognize deception: for example, to see a warning gesture in deceptive behaviour, without seeing it as a genuine warning. Seeing-in may thus have resulted from an arms race between deception and deception-detection.

1.5. Social behaviour and cognition.

1.5.1. Sterelny. Philosopher Kim Sterelny's complex argument (chapter 14) pursues the relationship between metarepresentation and social cognition within an evolutionary perspective on rationality. His argument falls into three broad stages. First, he describes the selective pressures that drive the genetic evolution of capacities for sophisticated information use by animals. Second, he describes the further distinctive pressures on hominid cognitive capacities associated with the social as well as genetic transmission of information. Finally, against the background of this dual view of the genetic and socio-cultural roles of cognitive capacities, Sterelny contrasts two conceptions of rationality that might be used to describe and contrast animals and human beings. The first is related to Kacelnik's biological rationality: a measure of overall cognitive efficiency, of the general evolved capacity of an agent to respond adaptively to the informational challenges it faces. Sterelny provides reasons for scepticism about this approach. He recommends instead a second, more specific conception of rationality in terms of metarepresentational folk logic. He views folk logic as an evolved capacity to assess socially (and especially linguistically) transmitted information, with its inherent openness to manipulation and high costs of error.

Sterelny explains, by reference to Peter Godfrey-Smith's views, how animals' capacities to use information are driven by the need to respond flexibly to variable environments. For example, in foraging, starlings must trade off the costs of declining efficiency in collecting food as their beaks fill up against the costs of flying back and forth to deliver their loads to waiting offspring. The efficient trade-off depends on variable information about the agent's environment. But adaptive behaviour also depends on a trade-off against the costs of gathering such information and of error: if the costs of information are too great, flexible responses reflecting information about varying environments may not be adaptive on balance. Some decision problems require environmental information that is relatively costless to acquire, but others require costly information. Environmental information can be degraded and become costly to collect in various ways that derive from competitive biological interactions. For example, other organisms pollute the informational environment by pre-empting the use of simple single cues for appropriate behaviour: prey hides, disguises itself, mimics poisonous organisms, and so on. And targeted agents often respond subversively to the actions of other agents, to block or thwart them: fast and frugal heuristics may work for catching a baseball, but are unlikely to intercept the non-ballistic, subversive path of fleeing prey. Such informationally hostile environments select for cognitive advances, such as the use of multiple cues rather than single cues to track features of the environment, and the decoupling of stimuli from automatic responses.

Hominid cognition reflects the selective pressures imposed by such informationally hostile biological interactions, as well as those deriving from environmental variation—from unstable climate and migration. The expansion of hominid populations into new environments created selective pressure for adaptive plasticity, as well as for specific adaptations by local groups to local environments. Like many other species, human beings construct important aspects of their own environmental niches. But human niche construction is distinctively cumulative, as one generation inherits the improvements of previous generations, such as cooking techniques and domesticated plants and animals, which in turn

induce profound changes in social environments and transform the risks and cognitive demands individuals face. Such changes occur faster than genetic adaptation can solve the specific problems they present, creating selective pressure to improve the cultural transmission of solutions between generations. In response, hominids have evolved distinctive genetic adaptations that facilitate social learning: capacities for imitative learning and teaching, an extended childhood for transmission of information between generations, language, and a division of cognitive labour. Once social learning is genetically enabled, however, it is self-amplifying through a process of cultural evolution. Runaway selection for the social transmission of information has, of course, costs as well as benefits: it is a powerful means of adapting swiftly to a changing environment, but it also risks extreme error costs, which in turn generates pressures for control of these risks.

Against this background of ideas about the evolution of cognition in general and of human cognition in particular, Sterelny sets out a contrast between two approaches to animal rationality. He argues that the evolution of rationality should not be viewed as the evolution of optimal cognitive design, and that is more fruitfully viewed as the evolution of folk logic.

The first approach conceives of rationality by reference to a general, optimally adaptive design for obtaining and using information. But Sterelny argues it is unlikely to be work. Why not? Because of a broader problem: there is no accepted, objective to measure overall fit between organism and environment or to assess the optimality of adaptations in general terms (as Lewontin argues). Fitness and adaptation are local, highly specific relations, between specific organisms and specific environments; definitions of general relations of fitness or adaptiveness have proved elusive. Moreover, explanatory work is done by specific first-order features of phenotype and environment, not by general fitness level. Sterelny argues that this problem with the idea of optimal adaptation carries across when it is applied to cognitive adaptations (see also and cf. Shettleworth 1998, 10, 570). Just as what counts as fitness enhancing is specific to organism and environment, so what counts as efficient use of information is specific to organism and environment. There is no global metric of cognitive efficiency. It is multi-dimensional and cannot be maximized in every direction simultaneously; the right trade-offs are sensitive to specific contexts. Good cognitive design in some contexts may be unsophisticated, as the costs of information may exceed its benefits. If we were to think of rationality in these terms, it would have no intrinsic link to degree of cognitive sophistication.

Sterelny favours the second approach, which conceives of rationality in terms of the more specific metarepresentational capacity of hominids to assess (as well as represent) the thought and talk of others. Such assessments express a folk logic: a set of more or less explicit norms about reasoning and the representations it produces. Folk logic presupposes metarepresentation, but involves more than that. Metarepresentation, of the signals of other agents as signals, might be useful in predicting their behaviour, even if it did not assess their accuracy. But folk logic assesses, as well as representing, representations and inferences. What is the function of such normative assessment? Folk logic is not required for inference; a capacity to make valid inferences does not require a capacity explicitly to assess inferences against norms. Do the norms of folk logic promote inferential efficiency? While they have done so in the history of science, folk logic presumably predates science. Like Proust,

Sterelny gives normativity a function related to the potential for social manipulation of information. Sterelny thinks that folk logic can be understood as an evolved response to the costs and benefits of using socially transmitted information. Social learning is risky as well as powerful. Socially transmitted ideas can exploit and damage their biological hosts, and put recipients at risk of manipulation. Some forms of social transmission are harder to manipulate than others: the social transmission of information through imita-ive learning of expert skills is harder to manipulate than the social transmission of information through lingui-stic signals, which are arbitrary and carry no intrinsic marks of reliability. As Sperber has suggested, folk logic filters out some of the bullshit from socially transmitted information, especially linguistically transmitted infor-mation. Sterelny argues that folk logic also responds to selection for cognitive plasticity in learning to use novel inferential techniques and representational media: to learn about reasoning flexibly, you need to be able to represent and evaluate your own reasoning. Thus, folk logic amplifies the effects of the social transmission of information, and responds to pressure for control of the risks of costly errors incurred by social transmission.

If the optimal cognitive design approach could be made to work, it might provide a graded conception of evolved rationality applicable in degrees across human and nonhuman animals. But as we've seen, Sterelny thinks it will not work. The folk logic approach he favours does not provide a conception of rationality applicable to nonhuman animals, since only human beings appear to have evolved folk logic. Nevertheless, it does provide a naturalistic view of how rationality develops from social intelligence. It also raises questions about genetic and cultural co-evolution: is folk logic, hence rationality, to be viewed as a genetically evolved form of control and guidance of cultural evolution, or as itself a product of cultural evolution—or a bit of both?

The chapters by Proust and by Sterelny place rationality in the context of the social intelligence hypothesis, that the cognitive demands of social life drive the evolution of intelligence.52 Social life and the social transmission of information create opportunities for animals to gain advantages over others by manipulating information, but also create opportunities to benefit by warranting the information they provide to others and by filtering the information provided by others. Rationality responds to the pressures to warrant and vet information that are generated by social learning. This general approach provides a plausible view of the evolution of rationality in the strong sense of human reasoning processes, but what implications does it have about prior stages of rationality in animals? Proust provides for intermediate stages of rationality in terms of metacognition without metarepresentation, while Sterelny regards rationality in the sense of metarepresentational folk logic as a hominid preserve.

1.5.2. Addessi and Visalberghi. Studies of Capuchin monkeys reported by primatologists Elsa Addessi and Elisabetta Visalberghi (chapter 15) provide a cautionary perspective on the link between social learning and rationality. Field observations of these monkeys, who forage in social in groups, make it tempting to suppose that they learn from observing others which foods are safe to eat by means of rational processes that would be natural to attribute to human beings in similar contexts. However, Addessi and Visalberghi's experiments

undermine the attribution of (what Kacelnik would call) PP-rational social learning to the monkeys, and instead suggest an explanation of their foraging behaviour in terms of lower-level processes operating in ways that are biased by social contexts, which nevertheless produce adaptive, or B-rational, behaviour in the relevant environments. This chapter thus draws together various themes already encountered, as well as that of the relations between social life and rationality, including: the methodological difficulties in attributing PP-rationality to animals, the possibility of explaining apparently rational behaviour in terms of lower-level rather than rational processes, the possibility of "outsourcing" to relevant environments some elements of the processes that support B-rational behaviour.

Social learning about food plays a large role in human eating habits, and many researchers have found it plausible to suppose that the feeding habits of other social primates are shaped by similar learning processes. The assumption has often been made that naïve animals observe experienced group members eating and learn from them which foods are good to eat and which are unsafe. Capuchin monkeys eat novel foods less than familiar foods, but eat more novel foods when in a group than when alone. Observations of social foraging in the wild that seem to support an attribution of rationally sophisticated social leaning are better explained, in light of experimental results, in terms of individual preferences and learned aversions, associative and trial and error learning, and social biases on individual learning.

Addessi and Visalberghi allow a monkey to see demonstrator monkeys eat a brightly coloured food, and the observer is offered a novel food of the same appearance and colour to eat, or a novel food of a very different colour. Whether the demonstrators are eating food of the same color or of a different color, the observer consumes more of its novel food when they can see demonstrators eating. That is, the observer eats no more of its novel food when its color matches that of the demonstrators' food than when its color differs. Interestingly, the observer consumes more of the novel food when there are more demonstrators eating. Other experiments showed that giving either the observers or the demonstrators a choice between two differently colored foods does not produce any tendency for the observers to choose the food that matches the food the demonstrators eat. Since eating novel foods is socially facilitated regardless of what is eaten and is not directed to specific novel foods, the monkeys do not appear to be learning what is safe to eat by observing what others eat. Nor does the experimental evidence support the hypothesis that monkeys learn which foods to avoid by observing which foods others avoid eating (absence is harder for animals to detect than presence).

What does appear to influence capuchins' choice among foods? The monkeys tend to prefer sweet foods and avoid bitter foods; sweetness reflects energy content and toxic foods are often bitter. They are cautious about novel foods; learning about foods begins with small samples, which are less likely to be lethal even if the food is toxic. If sickness results, the food is associated with it and avoided; if there are no bad consequences it is progressively included. These individual processes go a long way to enabling monkeys to select the right foods. However, social biases still influence an individual's food choices, though in a less 'targeted' and rationalistic way, in which social influences are mediated by an environment with certain properties. Social facilitation works non-specifically, by supporting the synchronisation of feeding activities rather than the learning of what to eat. However, since

individuals moving together are likely to encounter the same food sources in most environments, feeding synchronisation increases the probability that naïve individuals, along with knowledgeable individuals, eat the same foods. This social effect combines with individuals' preferences for certain tastes, their tendency to avoid novel foods, and food aversion learning, to increase the chances that monkeys feed in similar ways, which are biologically rational given their environments.

Thus, a widely assumed interpretation of wild foraging behaviour in terms of rational social learning does not stand up to experimental probing. Rather, the role of social facilitation in adaptive food choice can be understood in terms of lower-level processes and interactions with environmental features (compare the discussion by Papineau and Heyes of rival rational and associative accounts of imitative social learning in quails, this volume). The authors' methodological moral is that field observations may tempt us to make unwarranted attributions of human-like reasoning and, by themselves, can be an unreliable basis for attributions of process rationality, in the sense of Kacelnik's PP-rationality, to animals.

1.5.3. Connor and Mann. Cetologists Richard Connor and Janet Mann (chapter 16) are very aware of the difficulty of inferring from complex social behaviours in the wild to the cognitive processes that produce them. The alliance formation behaviours of bottlenose dolphins that their fieldwork meticulously documents are not only dauntingly complex, but appear to be intractable to experimental methods of probing process rationality. Nevertheless, their fascinating observations over many years of the social lives of over 600 bottlenose dolphins in Shark Bay provide a rich basis for theorizing about the biological rationality of dolphin behaviour. Connor and Mann compare the selective pressures imposed by social complexity in marine as opposed to terrestrial environments, and draw attention to the very different adaptations found in different cetacean species, which display variation in brain-to-body size ratios comparable to that between human beings and the great apes. Their work aims to contribute to our understanding of how the cognitive abilities demonstrated in captive bottlenose dolphins function in the wild, and of how the Machiavellian intelligence hypothesis–that the demands of social life select for advanced cognitive capacities–may apply to cetaceans.

Connor and Mann provide a review of dolphin social life, focussing on male alliances (both first-order and higher-order alliances), on female relationships, and on the roles of affiliative interactions, and making revealing comparisons with primates along the way. Like primates, dolphins mature slowly, and their offspring nurse for about four years; unlike primates, dolphins of both sexes maintain their natal territorial range as adults, so that from early life they can begin building social knowledge and negotiating relationships that may be important during their later reproductive years. And like chimps, dolphins live in a fission-fusion society, in which individuals associate in small groups of changing composition. However, while chimp social groups have strong boundaries, dolphin societies are open, with no boundaries between groups or territories but rather a pattern of overlapping home ranges. Primates are likely to know all the other individuals in their social group, but in Shark Bay dolphin A may know B and B know C, while A does not know C.

(a) Male multi-level alliances and social complexity. The strongest affiliations among the

dolphins are same-sex affiliations, though male-male and female-female affiliations differ in character. A distinctive and unusual feature of dolphin society is the existence of male multi-level alliances within a social group. Females with calves approaching weaning attract aggressive control or "herding" by first-order alliances of two or three males, which capture females from other alliances for herding and defend them against attack. Alliance partnerships may anticipate female fertility as existing calves mature, but may also function to test and consolidate male-male bonds. Stable first-order alliances in turn work together in second-order alliances that might have a dozen or more members; the more labile the first-order alliances, the larger the second-order alliances.

In general, alliances within social groups make for complexity, as when dolphin B may recruit A in alliance against C by means of affiliative behaviours such as grooming, while C is also trying to recruit A to ally against B. Such triadic interactions mediated by affiliative behaviour are prevalent between individuals within groups but not between groups, in non-human primates. However, social complexity increases dramatically if such triadic interactions are recapitulated between groups of allies as well, so that alliance ABC may compete with alliance GHI to recruit alliance DEF in a second-order alliance: consequences of individual interactions may be relevant at both levels of alliance. Such nested or hierarchical alliances generate a landscape of strategic options and high risks that exacerbate the demands on cognition and place a premium on social intelligence; among primates, they are characteristic only of human beings. They are also characteristic of dolphins.

Connor and Mann illustrate the challenges of nested alliance structures among male dolphins with several case studies: a 17-year history of the relations between dolphins Rea and Luc; a study of the 14-member second-order "super-alliance" WC; a study of the shifting multi-level alliances of three provisioned males given dead fish by human beings on daily beach visits. A striking observation of interactions among three second-order alliances PD, KS, and WC suggests a third level of alliance formation. While Connor and Mann's interpretation is cautious, it is tempting to describe the interaction as one in which an initial third-order alliance between PD and KS, which were not competing for their respective controlled females, was attacked by WC, which recruited PD to a new third order alliance that expelled KS, in which PD retained possession of its females but WC appropriated KS's female. However, it was difficult to determine whether PD actively "took sides" or merely remained neutral.

(b) Female associations and social learning. Females spend considerably less time than males do with their closest same-sex associations, and their group sizes vary widely; unlike male-male relationships, female-female relationships are very rarely antagonistic. Females may associate for reproductive functions, or to mob sharks. Members of one female group were observed to engage in petting the males in a trio that had been herding another female, while other members of the female group flanked the herded female, concealing her both visually and acoustically from the distracted males, and escorted her away to freedom—an unusual behaviour that may suggest intentional deception (at any rate, the males were fooled). The demands of foraging for single, mobile prey may account for aspects of female behaviour, such as neither stealing nor sharing prey (even with offspring). Moreover, females often specialize in specific, distinctive, foraging techniques, such as sponge-carrying (a form of tool-use) or beaching, which are passed on to their offspring who can be seen "practicing".

Sponge-carrying, for example, is found in less than 10% of the female population, and only in calves born to sponge-carrying mothers. Connor and Mann's work demonstrates that there is social transmission of several foraging "traditions" across generations, in addition to the amply-documented social learning of vocal activity.

(c) Affiliative interactions. Alliances among both primates and dolphins are mediated by affiliative interactions, such as grooming in primates or petting and synchronized swimming in dolphins, which may function to test the bond between individuals. In dolphins they are found not just between males in the same first-order alliance but also between males in different first-order alliances but the same second-order alliances. Males prefer to associate with other males from an early age, and homosexual social behaviours involving role exchanges and synchrony are likely to play a role in establishing male bonds. On one occasion males were observed to herd members of a rival second-order alliance as if they were females, directing at the rival males behaviours associated with consortship of females. To observers, emotions appear to play a role in aggressive and affiliative interactions among dolphins; they might function to guide complex social decision-making by providing a timely integration and comparison of the values of past interactions with others.

Connor and Mann end their discussion by reflecting on the selection pressures that could have led to high ratios of brain to body size in dolphins, given the large variation in brain-to-body size ratios among cetaceans. The energy-rich food of dolphins makes a large brain more affordable; but what benefits does it provide? Connor and Mann find unpersuasive the argument that large brains are needed to implement sophisticated echolocation functions for foraging, and support Herman's suggestion that social complexity drives the evolution of large brains in dolphins. Connor and Mann suggest that multi-level alliances among dolphins provide a link between theories of alliances and conflict within groups of nonhuman primates and theories of alliances and conflict—i.e. warfare—between groups of human beings. Warfare may increase the selective pressure for sophisticated social cognition because it requires individuals to cooperate against an enemy with other individuals in their group, with whom they are simultaneously in reproductive competition. Similarly, multi-level alliances demand that individuals base decisions at one level of interaction partly on the impact of those decisions at other levels. But what aspects of life in the open ocean provide the relevant trigger for such interdependence and social complexity? Perhaps cetaceans facing predation from sharks have nothing to hide themselves or their offspring behind except each other. Comparative study of the relationships among brain size, cognitive capacities, and behaviour among diverse cetacean species promises to increase our understanding of these and other issues about the role of social life in driving cognitive adaptations.

1.6. Mind reading and behaviour reading

The ability to understand the mental states of others, or mind reading is widely regarded as the most demanding form of social cognition. Mind reading was originally conceived as the exercise of a theoretical capacity for reasoning about the mental states of others, but simulation theories of mind reading are now prominent rivals to such "theory-theories". In recent years, two further significant developments have occurred in work on mind reading. First, a widespread view that mind reading is an exclusively human capacity has been

challenged, and controversy re-emerged about animal mind reading. Second (though relatedly), mind reading has been disaggregated into different aspects, some of which animals may have and others they may lack. For a time, tests for attributions of false beliefs to others were treated as the gold standard in assessments of mind reading. Children under about 4 years of age, and autistic persons, cannot answer correctly questions that require them to attribute to someone else a belief that they themselves believe to be false in order to predict the other's behaviour; instead, they attribute their own belief to the other. But recently, Tomasello (1999) and others have urged that a more graded view of mind reading capacities is needed, rather than an all-or-nothing assessment based on false belief tests (see also Povinelli and Vonk, this volume). There is evidence that the ability to attribute desires and intentions to others precedes the ability to attribute beliefs to others, both ontogenetically in children (Rakoczy et al, submitted), and phylogenetically (Tomasello 1999; Tomasello and Call 1997; Guzeldere et al 2002). At the same time, nonverbal experimental paradigms have been developed for assessing not just the capacity for false belief (see Tschudin's chapter), but also various other aspects of mind reading capacity. In particular, paradigms have been developed that focus on whether chimps understanding perceptual states of others, such as seeing, as opposed to epistemic states, such as beliefs.

An important aspect of the current controversy over animal mind reading concerns how results obtained using some of these nonverbal paradigms should be interpreted, and indeed how informative they are in principle. Here we represent the debate between two distinguished centres for the study of ape cognition, Michael Tomasello's group in Leipzig and Daniel Povinelli's group in Louisiana (see also Gomez 2004, ch. 8). While apes have been unable to pass standard nonverbal false belief tests, they have passed other nonverbal tests of mind reading for seeing, including those described by Tomasello and Call in their chapter. However, Povinelli and Vonk argue that success on such tests can just as well be explained in terms of behaviour reading as mind reading. We also represent inconclusive but suggestive data from nonverbal false belief paradigms applied to dolphins by Alain Tschudin, which casts further light on the difficult methodological issues in play.

In considering what kind of behavioral evidence favors animal mind reading as opposed to behavior reading relevant factors include not just methodological factors about the transfer of success to new tasks, the evidence base required for successful performance, and various appropriate controls. Another relevant factor is the conception of mental states that is in play. Given evidence may warrant attributions of mind reading on some conceptions of mental states but not on others. Crudely, some conceptions of mental states and their contents emphasize their internal, inferential connections among one another, while others emphasize their external connections, informational or teleological, to the world. Suppose an animal whose mind reading capacities are in question behaves in a way that suggests he attributes a given mental state to another animal but not a further, inferentially connected mental state. This may count against the mind reading interpretation less on an informational or teleological conception of the mental than on an inferential role conception of the mental.54 Differences over evidence for animal mind reading may reflect such implicit differences in the underlying conception of the mental. Moreover, the extent to which animal mind reading counts as an expression of social rationality in particular, as opposed to social perception or cognition more generally, may depend on the degree to which attributions of

mental states are viewed as carrying with them inferential commitments.

1.6.1. Tomasello and Call. Comparative psychologists Michael Tomasello and Josep Call (chapter 17) distinguish two approaches to interpreting the behaviour in non-human animals. Boosters interpret behaviour in psychologically rich ways; scoffers prefer psychologically lean interpretations. The ultimate boosters think there are no significant between human and nonhuman cognition, while the ultimate scoffers are radical behaviourists, who do not find it useful to talk of cognitive processes at all. Boosters about primate social cognition attribute understanding of others' psychological states to apes; scoffers regard apes as clever behaviourists themselves, who merely read and respond to others' behaviour but lack understanding of others' psychological states. In their chapter, Tomasello and Call compare richer and leaner interpretations of recent data from four experimental paradigms concerning whether chimpanzees know what others can or cannot see, or only what others are looking at. They argue that the booster hypothesis, that apes know what others see, is better supported than the various scoffer hypotheses that would be needed to explain these results.

(a) Gaze following. Experiments show that chimpanzees reliably follow the gaze of other chimpanzees and human beings, even when doing so requires looking past and ignoring other novel objects along the way, or moving in order to follow a gaze to a target behind a barrier. If they find nothing of interest, they check back and track the other's gaze again, and eventually stop following the gaze of individuals who repeatedly look to locations where no salient target is found. The booster interpretation of this behaviour is that chimpanzees follow gaze because they want to see what the other sees. The scoffer interpretation avoids attributing to the chimpanzee knowledge that the other is seeing something and instead attempts to explain the behaviour by appeal to biological predispositions and learning: perhaps they are disposed to orient the way others orient, or have learned to look in the direction others are oriented.

(b) Competing for food. A series of experiments place a subordinate and a dominant chimpanzee in rooms on opposite sides of a baiting area in which food is placed in various positions in relations to barriers before the chimps are allowed to go for it. The dominant chimp will take any food she can see, but the subordinate can monitor the dominant's visual access to the food. It has been found that subordinates chimpanzees go for food that a dominant chimpanzee cannot see because it is hidden from view behind a barrier much more often than they go for food that the dominant can see.

Tomasello and Call consider and dismiss five "scoffer" interpretations of this behaviour, some of which are ruled out by further experiments. The subordinates are not merely responding to the initial behaviour of the dominants, since when they are given a head start they still go for the food the dominant cannot see. The subordinates do not merely view the barrier as something that might slow down the dominant's physical approach to the food, since they do not favour food placed behind transparent barriers. The subordinates do not simply prefer to forage near barriers rather than in the open, since when all the food is placed near barriers they still prefer food dominants have not seen. The subordinates do not simply respond to the presence or absence of a dominant during baiting, since they go freely for food seen by one dominant if the first dominant is replaced by a second who has seen nothing;

they seem to know what particular individuals have seen or not seen. Finally, the subordinates do not simply avoid food that a dominant has looked at; if the subordinate monitors a dominant watching the food being placed initially, but the food is then moved to a new location in view of the subordinate but not the dominant, the subordinates still go for it even though the dominant has looked at it. It is not plausible to suppose that the chimpanzees have had past learning experi-ences involving transparent barriers, relocated food, and so on, that would explain these results. A better interpretation is that chimpanzees do not just follow the gaze of others, but also know something about what others see, on the basis of which they infer what others are likely to do and hence decide what they themselves should do.

(c) Begging and gesturing. In these experiments, chimpanzees can beg for food from one of two trainers in different attentional states. The chimps chose to beg from a trainer facing them rather than one turned away from them. But they did not beg more from a trainer wearing a blindfold over his mouth as opposed to his eyes, or from one with a bucket on his shoulder as opposed to over his head, or from one with hands covering his ears as opposed to his eyes, or from one with eyes open as opposed to closed, or from one turned away but looking over his back at the subject as opposed to turned fully away. The scoffer interpre-tation of these results is that chimps attend to the body orientation of others, but know little if anything about what others see. However, in a modified design in which chimps faced only one human trainer rather than having to chose between them, they gestured differently depen-ding on where the trainer was looking and the orientation of her face (see the comments by Povinelli and Vonk on these results, in section 3 of the appendix to their chapter, this volume). A booster interpretation of these results is that body orientation indicates the trainer's disposition to give the chimp food, while face orientation carries information about whether the trainer can see the chimp's begging gesture. Another line of experiments, in which chimps compete with human trainers for food, supports the view that face orientation gives chimps information about what others can see. When a trainer's body is facing one piece of food but his head is turned away toward another piece of food, chimps attempt to steal the food in front of his body rather than the food the trainer can see. A scoffer interpretation of these and other competition-with-a-human-being results is that chimps do not understand what someone can see, but simply avoid approaching food if they can see a competitor's face. This hypothesis was ruled out by an experiment in which chimps avoided going for food when the chimp could not see the human competitor's face but it was clear that the human competitor could nevertheless see the chimp approaching the food.

(d) Self-knowledge. The last of the four paradigms is one also discussed by Call's chapter on metacognition, in which one of two tubes is baited; to get the reward, the chimps have to touch the baited tube. Recall that chimps looked into the tubes before choosing which to touch more often when the tubes had been baited out of their view than when they had watched the baiting. A scoffer explanation of these and related results could appeal to hypothetical past learning of a difficult conditional discrimination. The alternative booster interpretation is simply that chimps know what they themselves have and have not seen, and hence know to look in the tubes when they have not seen them being baited.

Taken together, Tomasello and Call argue, these results make a strong case for the booster

hypothesis that chimps understand what others (and they themselves) can and cannot see. The alternative scoffer hypothesis requires a dozen different hypotheses to explain the same phenomena, including various ad hoc and unlikely prior learning scenarios for which there is no independent evidence. They aren't impressed by either parsimony or Morgan's Canon as a reason to favour the scoffer view. It isn't clear whether parsimony here favours invoking fewer different hypotheses, as the boosters, do, or invoking only behaviouristic learning rather than mind reading, as the scoffers do. Morgan's Canon counsels us to avoid postulating higher-level mechanisms to do explanatory work that can be done by lower level mechanisms, but the higher-lower distinction is problematic. Finally, learning is not incompatible with understanding seeing: they hypothesize that chimps may learn about what others can see more readily than they learn arbitrary associations, just as they can learn causal relations more readily than arbitrary associations (see Call's chapter for discussion). While Tomasello and Call believe that the evidence that chimps understand seeing is now overwhelming, they caution that this does not mean that the chimps also understand other psychological states. Mind reading is not all-or-nothing; chimps may be able to understand some psychological states but not others, and empirical evidence is needed for each case.

1.6.2. Povinelli and Vonk. Comparative psychologists Daniel Povinelli and Jennifer Vonk (chapter 18) mount a methodological argument that there is no compelling evidence for the claim that chimps understanding seeing in others because the kinds of experimental paradigms that have been used in support of this claim, such as those discussed by Tomasello and Call in their chapter, are in principle incapable of providing evidence for it. A logical problem, they argue, precludes obtaining empirical evidence for animal mind reading by employing such paradigms; until the logical problem is resolved, the empirical issue of ape mind reading is not joined.

(a) Louisiana vs. Leipzig: The logical problem with existing experimental paradigms for animal mind reading. They explain the logical problem as follows. The standard experimental aim is to design an experiment in which a capacity to represent specific behaviours and behavioural invariances in relation to the environment predicts one response, while a capacity to represent mental states in addition predicts a different response. This assumes that it is clear, on folk psychological general principles, that a certain response would only be generated given the additional capacity to represent mental states. But this assumption is unjustified. The responses typically taken as evidence for representations of mental states are just as well explained in terms of representations of behaviour and environment. The animal's representations of behaviour and environment would in these cases have to provide the evidence for its inferences to "intervening variables", that is, to others' mental states. An animal cannot avoid the work of detecting invariances and regularities and nuances in others' behaviour by attributing mental states to them, since the former provide evidence of the latter. So what explanatory power is gained by attributing to the animal representations of mental states in addition to representations of behaviour? Mind reading presupposes smart behaviour reading in these cases, we might say, and there is no evident explanatory work being done by mind reading that smart behaviour reading isn't already doing. Attributions of mind reading in addition therefore appear to be explanatorily redundant. For example, in the begging experiments, a chimp's attributions of seeing to the

experimenter must be based on the chimp's representations of the experimenter's body and behaviour, e.g. on whether her eyes are blindfolded or not. That is, the behavioural representations on which mind reading depends must be present in any case. But then why can't the chimp predict how the experimenter will respond to a begging gesture on the basis of such behaviour reading by itself just as well as on the basis of such behaviour reading plus attributions of seeing to the experimenter? And if the chimp can predict the experimenter's behavior equally well by means of behaviour reading as by means of behaviour reading plus mind reading, then the chimp's own behaviour in such experiments is equally well predicted by attributing behaviour reading to it as by attributing behaviour reading plus mind reading. The chimp's behavior thus provides no evidence for mind reading; mind reading has no added explanatory value. More generally, the experimental paradigms currently in use cannot distinguish between behaviour reading by itself and behaviour reading plus mind reading.

Povinelli and Vonk consider and reject the idea that explanations of behaviour that appeal to mind reading are more parsimonious than explanations that appeal to representations of many different, specific behavioural regularities. Their reason is that mind reading by chimps in these experimental paradigms presupposes that chimps represent the relevant behavioural regularities in any case; there is no increase in parsimony by adding representations of mental states as well. Adding control conditions to rule out reliance on representations of a specific behavioural regularity, as the Leipzig group has done, does not address the point that mind reading in these paradigms must rely on representations of some behavioural regularity or other, which are really doing the work. This is a logical rather than an empirical problem.

(b) How to avoid the logical problem. To avoid the logical problem requires finding a category of social behaviour that cannot be explained by attributing representations of behaviour but that can be explained by attributing representations of mental states. There are two broad approaches here (cf. Sterelny 2003, 75).

First, one can focus on the output side: on the functions of representations of mental states. Perhaps it can be shown that there are limits to the functions that representations of behaviour can have in generating social complexity, but that representations of mental states can have further functions. If so, then evidence from animals that such further functions are being performed would be evidence for representations of mental states in addition to representations of behaviour. Although Povinelli and Vonk do not pursue this approach, arguments exist by means of which it might be developed. For example, it can be argued that "mirror metaheuristics" provide cooperative solutions to certain games, such as one-off Prisoner's Dilemmas, that are not made available by accurate predictions of behaviour alone but are only available to those who can mind read as well as behaviour read.55 On this view, representations of others' heuristics have further functions than do representations of their behaviour, even if the latter provide evidence for the former. If there were evidence that animals can cooperate effectively in one-off Prisoner's Dilemmas (in which they do not repeatedly play against the same partners—and of course in which solutions are not genetically determined), that might provide evidence that they were doing more than predicting one another's behaviour and were indeed representing the heuristics of others and their similarity or dissimilarity to their own heuristics.

Povinelli and Vonk develop a second way of avoiding the logical problem, which focuses in effect on the input side: on the evidence for representations of mental states. The aim here is to find tasks that avoid the crucial feature of the logical problem—that is, tasks for which there is no plausible representation of the others' behaviour that is presupposed by the postulated representations of their mental states. The desired tasks should be such that success can be explained by attributing representations of others' mental states, but cannot be explained by attributing representations of others' behaviour–because no relevant representations of others' behaviour are available. Following a lead provided by Cecilia Heyes (1998), Povinelli and Vonk propose pursuing a class of such tasks for which successful performance can only be generated by mapping one's own experience onto that of others: self-other inference tasks. Such inferences are familiar to philosophers who study the problem of other minds under the heading of "arguments from analogy".

Povinelli and Vonk put forward an example of an experimental task of a conceptually different nature from those they criticize, which avoids the logical problem and hence succeeds in joining an empirical issue about ape mind-reading. Chimps could be given first-person experience of wearing two buckets containing visors. The buckets are different colours; while the visors look the same from the outside, one can be seen through while the other is opaque. When two experimenters don the buckets, will chimps who have had first-person experience of wearing the buckets prefer from the first trial to beg from the experimenter wearing the bucket containing the see-through visor? If chimps represent only others' behaviour, the predicted answer is "no": since they have never seen anyone wearing either bucket before, no information is available to them about what others wearing buckets do, on which to base such a preference. If chimps can represent others' minds as well, the predicted answer is "yes" since if they can represent their own experience and then attribute an analogue of it to the other, information is available to them on which to base such a preference.56

Povinelli and Vonk emphasize that the disagreement between Louisiana and Leipzig does not turn on whether mind reading is an all or nothing capacity; both sides agree that it should not be assumed to be all or nothing. Indeed, this view is reflected in the Louisiana group's focus on perception rather than belief in its seeing/not-seeing studies with chimps..

This chapter directs its attention to recent nonverbal paradigms for assessing the attribution of seeing to others, given that claims for ape mind reading have been made based on that paradigm. However, the logical challenge laid down is quite general, and in principle also applies to nonverbal false belief paradigms. While the consensus is that apes fail nonverbal false belief tests, the evidence is as yet inconclusive for dolphins.

1.6.3. Tschudin. Comparative psychologist Alain Tschudin reviews his work using nonverbal false-belief paradigms with dolphins, which has yielded promising but equivocal results. He illuminates various method-ological challenges such work faces and how they can be met, and makes a strong case for the importance of further research with these highly social, cognitively sophisticated animals to resolve the false belief issue.

Tschudin places the question whether dolphins can attribute beliefs to others in the context of both (a) the hypothesis that social complexity drives the evolution of advanced cognitive capacities and (b) the correlations between social complexity and social group size with brain-to-body ratios in both primates and dolphins. The background features of social complexity and large brains in relation to body size that have prompted investigation of the mind-reading capacities of apes also apply to dolphins. Moreover, dolphins exhibit some of the precursors to mind reading, such as a capacity for joint attention, spontaneous comprehension of referential pointing, and gaze following.

(a) Pilot study. Given these supporting considerations, Tschudin and colleagues conducted a pilot study with dolphins using a hider-communicator false-belief paradigm. As he explains, it is easier to distinguish behaviour expressing an agent's own belief from behaviour that expresses the agent's attribution of a belief to another if the other-attributed belief is inconsistent with the agent's own beliefs or "false". In the pilot study, the dolphin can see that a hider baits one of two opaque boxes with a fish, but cannot see which one. However, the dolphin can also see that a different trainer, a communicator, can see which box has been baited. After baiting, the trainer tapped on one of the boxes, and the dolphin indicated its choice of one of the boxes and was given its contents: a fish, if it indicated the box containing the fish. During the training trials, the boxes were not switched after baiting, so that the dolphin learned to obtain fish by choosing the box on which the communicator tapped. During the test trials, the communicator departed after baiting, and the dolphin observed the other experimenter switch the boxes. The communicator then returned, and tapped the "wrong" box. All the dolphins passed this false belief task.

However, there were some methodological problems with the design of the pilot study, which tempered excitement about the result. First, a confound may have been generated by pre-test control trials in which the boxes were switched after the tap by the communicator while the dolphin watched, and in which the fish was moved from one box to the other after the tap while the dolphin watched. The dolphins passed these tests, which were intended to ensure that the dolphins understood spatial displacement and could ignore the communicator when appropriate. However, they may have learned from these trials simply to choose the opposite box to the tapped box whenever there is a switch. Work with four-year-old children underscores this concern. Second, the same experimenter baited the boxes and presented the dolphin with a choice between them, and may inadvertently have cued the correct response to the dolphin. Subsequent experiments were designed to address these problems.

(b) Experiment 1. In experiment 1, the pre-test controls were omitted and true-belief tasks (the boxes are switched in the presence of the communicator) were interspersed with false-belief tasks to prevent reversal learning. Moreover, different experimenters were used for baiting and presenting. One of four dolphins failed to pass the training phase, perseverating to the left; but the other three succeeded on the first false-belief trial and two of these succeeded on the first true-belief trial. Overall, there was a significant relationship between whether the communicator's indicated "beliefs" were true or false and whether the dolphins responded in accord with them or not. But when the results were pooled on a given trial for all animals, or for a given animal on all trials, they are of equivocal significance; this may be an effect of the small number of trials. As well as the need for an increased sample size,

further methodological concerns arose from this experiment, including a worry that the communicator herself may have inadvertently provided cues other than the tap signal, and that the dolphins, who had also been used in the pilot study, might remember a reversal strategy learned during the pilot study.

(c) Experiment 2. Experiment 2 was designed to address these concerns, using a naïve animal and interspersing trials in which the communicator did not know whether the boxes had in fact been switched during her absence or not. Cueing and learning were now thoroughly controlled for. Unfortunately, the naïve dolphin did not pass the training phase but perseverated in various ways, so the experiment was discontinued. This dolphin did not fail the false belief test; rather, he never met the training conditions for taking the test.

The combined results from the pilot study and experiment 1 may make it tempting to ascribe mind reading to dolphins. But this would be premature; alternative explanations need to be ruled out. A reversal learning explanation is possible, as indicated; a larger sample size and naïve dolphins are needed to assess whether dolphins can succeed on first trials with all appropriate controls in place, which reversal learning could not explain but mind reading could. Another alternative explanation would be one of the kind Povinelli and Vonk press, namely, that the dolphins attribute a behavioural state concerning visual access rather than a mental state to the communicator, and base their responses on this. Tschudin argues that a behavioural rule could not account for success on first false belief trials, because it would have to be learned; first trial performance should be at chance in the absence of mind reading. (However, Povinelli and Vonk might reply that mind reading must be based on behaviour reading even in first trials, and that the added value of mind reading in explaining first trial success is unclear.) Another explanation is in terms of inadvertent cueing of correct responses by experim-enters; unfortunately, in experiment 2, with the best design to exclude various forms of cuing, was discontinued.

A mind reading interpretation gains support from the background considerations already reviewed, though it would have to explain the failure of two dolphins to pass the training phase. These failures may simply reflect differences between individual animals. Moreover, the naïve dolphin used in experiment 2 was socially isolated in relation to other dolphins, which may have interfered with the development of his normal social capacities despite his interaction with human trainers. To provide points of comparison, experiment 2 was also run with seals and children. The seals and four-year-old children failed, and only 3 out of 14 five-year-olds passed false-belief trials on first trial. Moreover, children have been attributed the capacity to mind read based on experim-ental designs similar to that of the pilot study, with fewer controls than used in experiment 1 with dolphins. Arguably, evidence for mind reading by dolphins is being held to higher standards than for children.

Given the methodological lessons of Tschudin's work on dolphin mind reading, the supportive direction of his results and lack of clear failure by dolphins on false belief tests, further research is warranted on the question of mind reading by dolphins and, more generally, on the social cognitive capacities that support their complex social behaviour. If dolphins were to be successful on false belief tasks with all appropriate controls in place, this would raise further questions: in particular, whether success would be equally open to

explanations in terms of behavior reading as in terms of mind reading; and, if the latter, what further experiments might reveal whether such success were best understood as an expression of theoretical rationality (as in theory-theory accounts of mind reading) or of practical rationality (as in simulation theory accounts).

1.7. Behaviour and cognition in symbolic environments

1.7.1. Herman. Cetologist and psychologist Louis Herman (chapter 20) is a founding pioneer of the experimental study of dolphin cognition. Without his work we would have far less understanding than we do of the remarkable intelligence of these animals. His chapter reviews the cognitive accomplishments of his four dolphins, Akeakamai, Phoenix, Elele, and Hiapo, in declarative, procedural, social, and self-related domains, presenting evidence for their rational responses in these domains. Throughout his discussion the dolphins' trained facility in understanding symbolic gestures is apparent, raising questions about how this capacity to understand symbolic gestures is related to the rational responses dolphins display.

A number of factors make dolphins compelling subjects for experimental studies of cognition. Their ratio of brain to body mass is second only to that of human beings; they live to 40 or 50 years and have a protracted period of development; they are highly sociable and naturally inhabit highly complex social environments. While their echolocation abilities are extremely well developed, these do not explain their large brains, which are more plausibly viewed as subserving their intelligence in a variety of domains. In Herman's view, intelligence is manifested in behavioural flexibility, the ability to adapt behaviour to a changed environment to enable effective functioning, in ways that go beyond biologically programmed and learned behaviours. The capacity for intelligently flexible behaviour can be present to various degrees in various dimensions, and provides the foundation for rational behaviour. A rational animal can perceive the structure and function of the world it occupies, and can make inferences that enable it to function effectively, based on its model of that world; moreover, it can incorporate evidence to model a changing world, or even radically different worlds, and adapt its behaviour appropriately, flexibly generalizing, inferring and innovating. Wild dolphins are able to respond flexibly and to function effectively in laboratory settings that are radically different from their natural environments, in ways that provide evidence that dolphins have created accurate models of their new world.

(a) The declarative and procedural domains. The declarative domain concerns abilities to perceive things and their properties, and to understand references to these things; the procedural domain concerns understanding of how to do things and of how things work. Herman discusses these domains together, since some of his experimental paradigms require both declarative and procedural knowledge. Various paradigms provide evidence of rational behaviour in these domains.

Akeakamai responds correctly to communications that use a symbolic language of hand gestures with an inverse grammar: the order of the symbols is not the order in which responses to them are required, and the significance of symbol order depends on the whole sequence. For example, the symbol order indirect object/direct object/relational action requires the dolphin to perform a particular action on a particular object in relation to another

object. Swimmer/surfboard/fetch requires the dolphin to bring the surfboard to the swimmer, while surfboard/swimmer/fetch requires the dolphin to bring the swimmer to the surfboard. Akeakamai responds correctly to novel sequences using this relational syntactic frame, treating objects grammatically correctly as direct or indirect objects according to their place in the order of symbols. Novel sequences that are semantically anomalous, requiring impossible performances, produce no response other than continued attention to the experimenter. In other cases, Ake takes initiatives to enable instructions to be complied with. Innovating in response to the instructions to jump over a surfboard that was resting on the side of the pool, Ake first moved the surfboard into the pool and then jumped over it; asked to put a ball in a basket that was already in a basket, she first removed it and then replaced it. Without specific training, she typically decomposes syntactic anomalies and acts on a syntactically correct part of the string. When modifiers left and right trained using non-relational commands, such as swim through left hoop, are added to the inverse relational syntax distinguishing direct and indirect objects to create a five-place syntax, Ake immediately, with no further training, generalizes and responds correctly. For example, she responds correctly to put the right ball in the left basket and to put the left ball in the right basket. She responds correctly (on over 80% of trials) to queries referring to absent objects, using "yes" and "no" paddles to questions about whether certain objects are in the pool after observing objects being thrown over her head into the pool behind her. If asked to transport an absent object, she spontaneously, without training, pressed the "no" paddle. If asked to transport a present object to an absent object, from first trial and with no training, she pushed the present object to the "no" paddle, and has continued to respond in this way for any combination of present direct objects and absent indirect objects. While enculturated chimps have shown little initial interest in televised images, Ake responded immediately and with no TV-specific training to televised images of trainers giving gestural commands, even when these images were degraded to include only the hands; and she outperformed some human staff members in responding correctly to still further degraded images. With no TV-specific training, another dolphin Elele generalized to perform match-to-sample tasks from televised images of samples, as accurately as from real samples. All four dolphins generate novel, self-selected behaviours during training sessions if and only if they are given a create command.

(b) The social domain. Evidence of flexible generalization, inference, and innovation by dolphins extends to the social realm. In human beings, pointing is a means of sharing attention to the same object with another. Unlike chimps, Ake spontaneously understands the object of referential pointing by human beings, including cross-body pointing, responding correctly to instructions accompanied by points. Remarkably, without training she integrated points into the relational syntax she had learned, generalizing the inverse grammatical rule to the order of points. Herman suggests that dolphins may understand pointing by generalization from an analogue of pointing provided by their highly focussed echolocation beams. These may provide a natural mechanism for sharing attention; it has been shown that dolphins can tell what object another dolphin is echolocating.

As Connor and Mann explain in their chapter, synchronized swimming is a natural affiliative behaviour. All Herman's dolphins have been trained to respond in synchrony to commands preceded by a tandem gesture, and they generalize this command from the few behaviours they are trained on to others. Most remarkably, the first time they are exposed to the

combination of tandem and create, pairs of dolphins respond correctly. Typically, they swim side by side for longer than they would if asked to synchronize on a specific behaviour, and then execute a selected behaviour synchronously, such as a spinning leap while spitting water from their mouths (which requires extra water to be taken into their mouths before they leap). This creative coordination is not taught, and Herman has not been able to determine how it is accomplished; close video analysis does not reveal a leader and a mimic. However, dolphins are talented vocal and motor mimics, and also show creative generalization here. Ake was trained to mimic vocally a variety of electronically produced sounds of different waveforms, and when the models were out of her preferred vocal range she spontaneously reproduced the sound contour at an octave above or below the model. She also recognized tunes transposed by an octave; octave generalization is a rare phenomenon among animals. The dolphins respond to a gestural command mimic in combination with motor behaviours modelled by human beings as well as by other dolphins. When mimicking human beings, the dolphins spontaneously analogise between human and dolphin body parts, raising a tail, say, when a human being raises a leg; and they generalized their mimicry to televised models.

(c) The domain of self-knowledge. As well as displaying mirror self-recognition, dolphins display awareness of their own behaviours and body parts. Phoenix repeats her past behaviours or chooses a different behaviour on request, requiring her to represent her past behaviour and generate a match or a mismatch. Elele associates her various body parts with gestural symbols, and can respond correctly to instructions requiring her to use different body parts in the same way or the same body part in different ways; she can respond to such requests involving novel combinations not in her natural or learned repertoire on the first occasion of presentation.

These and other paradigms that Herman reviews provide multiple instances of rational responses not specifically trained and requiring flexible generalization, inference, and innovation. Correct responses are inferred without specific training when new semantic elements are inserted into familiar syntactic slots, when the syntactic positions of objects are reversed, when novel syntactic structures combine previously familiar syntactic structures, including references to absent objects, when symbolic gestures are represented in novel televised formats, when pointing gestures are incorporated into familiar syntactic slots, when responses require creative coordination with another dolphin or untaught analogies between human and dolphin body parts or generalization across octaves. Initiatives are taken to rearrange objects so that instructions can be complied with. This level and variety of flexible generalizations and innovations suggests not merely “islands of practical rationality”, in Hurley’s terms, but significant conceptual and inferential abilities across several domains.

What processes select for evolution of this level of intelligence? Herman makes the suggestion, which Connor and Mann, and Tschudin also find plausible, that the pressures of social complexity play a key role in driving the evolution of dolphin intelligence. Herman suggests that these operate along with various ecological pressures. However, Herman’s work raises further compelling questions about the role of symbols in his dolphins’ remarkable abilities: does training with symbols support or extend dolphin cognitive abilities, does it enable dolphin rationality, does it express an independently grounded social intelligence, or does it reflect natural skills at symbolic communication that we have not yet

been able to evidence in the wild?

1.7.2. Pepperberg. Related questions about the role of symbol training in facilitating remarkable cognitive performances by animals are raised by comparative psychologist Irene Pepperberg's work with African Grey parrots (chapter 21). Pepperberg draws on her work to probe the relationship between animal intelligence and rationality, and between standards of rationality appropriate to human and to animal behaviour. Behaviour that is intelligent, in the sense that it displays sophisticated cognitive abilities that are biologically adaptive, may not count as rational by human standards.

In some cases biological intelligence converges with human standards of rationality, as when the adaptive mating behaviour of female great tits conforms to transitive inferences about the rank of various males. The intelligent behaviour of Pepperberg's parrots also often converges with human standards of rationality. Alex's accomplishments include comprehending and producing English labels for objects, for specific colours, shapes, and numbers, for the categories of material, colour, and shape, and various English commands, requests, and responses. Using these skills, he responds to many tasks in ways that would be regarded as rational in a human child. He produces correct responses in English to recursive, conjunctive questions in English that combine labels in new ways; he correctly answers questions about size, quantity, colour, similarity or difference of attributes in application to new objects; he requests absent objects. For example, he can label the quantity of a specific subset of objects in a complex heterogeneous array: when shown a tray containing 1 blue box, 3 green boxes, 4 blue cups, and 6 green cups, and asked how many green cups there are, he will respond by saying "six". This capacity generalizes to other questions about the same objects and to new objects, such as a question about what shape the green paper is, when shown an array of 7 objects of different colours, shapes, materials.

However, Pepperberg here focuses not on these well-known accomplishments, but on further labelling behaviours outside the usual training process and on certain "mistakes" the birds make, for what they reveal about the relationships between biological intelligence and various standards of rationality.

(a) Transitions to referential labelling and model rival training. Pepperberg has developed a distinctive training process, the model-rival system. The parrot eavesdrops on two trainers talking about interesting objects (compare Kanzi's eavesdropping on attempts to language-train his mother Matata; see Savage-Rumbaugh, Rumbaugh, and Fields, this volume). One trainer asks the other questions and rewards correct responses by giving the other the labelled object(s), as well as by praise. Errors are also demonstrated and punished with scolding and withdrawal of the object(s). The second trainer is the parrot's model as well as its rival for attention. The trainers reverse roles, and also begin to include the parrot in their interactions. Pepperberg has demonstrated with juvenile parrots that training is unsuccessful if it omits any element of this three-way social modelling process. Moreover, when the birds are given model-rival training for some labels and other training techniques are used with other labels, the birds "practice" the former labels but not the latter to themselves after working hours.

Pepperberg compares early label learning by her parrots with that by children. Label

learning is initially slow and difficult for both, but picks up speed; and considerable self-initiated learning appears to occur outside of training for the parrots, as when children play naming games. For parrots as for children, she argues, first labels are qualitatively different from later labels; they are not fully representational, referring to immediately present items and often overextended or underextended, perhaps reflecting lack of memory capacity or capacity for categorical classification. In parrots as in children, she suggests, the transition to representational use of labels parallels the transition from a self-centred orientation to recognition of others as information sources and to the rudiments of social rationality, a transition mediated by the cognitive functions of emotions. Underlying neural developments, especially in mirror systems in human beings and in possible analogous structures in avian brains, may contribute to these transitions by enabling the combinatorial recreation of complex observed structures of speech and action.

The parrots' transition from self-centred early label learning is seen, for example, in sound play outside training sessions, in which they spontaneously recombine or vary labels to produce new sounds: after learning "grey", Alex produced "grape", "grate", "grain", "chain", and "cane". If such new sounds are rewarded with corresponding objects, they are readily incorporated into the bird's repertoire of labels. A label learned in a specific context, such as "wool" for a woollen pompon, might be used while pulling at a trainer's sweater. It is possible that by means of these behaviours the birds, in the early stages of forming categories and representations, are rationally probing their trainers for information about the referents of sounds and testing hypotheses. The model-rival technique may facilitate the transition to representational use of labels by using two trainers, to separate the utterance from its referent or the required response, and by providing a model to imitate that generalizes across role reversals. Model-rival training has produced significant improvement in the communicative and social abilities of autistic children; Pepperberg suggests that it engenders exceptional learning, which would not occur given normal input.

(b) Responses to being tricked, and playing tricks. The parrots demonstrate an appreciation of object permanence by retrieving an object that has been covered by a container and then passed behind various screens. When the birds have witnessed a particularly desirable object being covered and the object they find is not the expected one (the experimenter has "tricked" them by an unseen swap), the birds do not continue looking for the desirable object, but rather display apparent surprise and anger. While their emotional response to not finding the expected object expresses an intelligent awareness, it is not clear whether this response is more or less rational than continuing to look for the desired object would be.

In another case, a parrot appears to play a trick on the experimenter. When shown a tray of seven items of different colours, shapes and materials, the parrot may be asked, for example, the colour of the square wooden object. He has already demonstrated mastery of this type of task; but on occasion, instead of providing the correct answer, he provides each of six possible wrong answers, and then repeats them. He thereby sacrifices a reward, but also produces a frustrated emotional response from his trainer. No doubt considerable intelligence is needed to perform in this systematically perverse way. But is his behaviour irrational? Or is it a rational means of producing an emotional response from his trainer that he enjoys more than he would the reward?

Pepperberg concludes by emphasizing the need to find experimental methods of assessing rationality across various taxa that are sensitive to the possibilities of continuities and convergence in rational functioning, but do not rely inappropriately on human-centred values and goal assumptions.

1.7.3. Boysen. Primatologist Sarah Boysen (chapter 22) directly addresses the role of symbol training in facilitating the cognitive performance of enculturated chimps. She argues that immersion in symbol-laden human culture and long-term social relationships with human beings dramatically increases their attentional resources. By permitting an animal to process selected information about objects, she suggests, symbols can minimize interference from low-level dispositions and enable them to respond rationally. Here she surveys work from three paradigms: on the way numerals modulate chimps' evaluative dispositions, on the way numerals and symbols for "same" and "different" modulate reaction time, and on chimps' use of scale models to solve problems.

(a) Candies and numerals: interference effects and symbolic facilitation. Boysen explains that, in certain circumstances, multiple evaluative dispositions (which perhaps function at different levels in the cognitive structure of the animal) may be expressed, and that these may conflict. Such conflict is illustrated in her work on chimpanzees' capacity to make quantity judgements. This work showed that chimps were unable to perform in an instrumentally rational way, that is, to point to a smaller quantity of candy in order to obtain a larger quantity; moreover, their performance was poorest when the disparity in quantity of candies was largest, so that they stood the most to gain. However, these chimps had previously been trained to use numerals, and they succeeded immediately on the task when candies were replaced with numerals: they pointed to the smaller of two numerals and were accordingly rewarded with a number of candies corresponding to the larger numeral. Related results have been obtained with children, who respond more rationally to photos of food than to actual food. Boysen suggests that the inherent incentive features of the candy introduce a bias toward the larger quantity that conflicts with an instrumental associative disposition based on the reinforcement contingency (you get the larger quantity if you point to the smaller quantity), which they appeared to have learned. (When different quantities of rocks are substituted for candies, the chimps also respond incorrectly, suggesting that an interference effect can also be based on perceived array mass.) Numerals represent quantity while abstracting from the interfering incentive properties, thus modulating the interaction of the chimps' conflicting dispositions.

To assess whether the presence of numerals fostered an abstract mode of processing that would overcome the incentive effect of the candies, Boysen next conducted a feasibility study in which chimps were presented with mixed arrays of candies and numerals on the same reversed reinforcement contingency. In the feasibility study, the chimps did not simply point to the candies in the mixed array sessions, and were only slightly worse at pointing to the smaller quantity in mixed arrays than in numeral-numeral arrays. The mixed array feasibility study thus supported the hypothesis that numerals facilitate the instrumentally rational response.

However, in a full-scale study that presented chimps with candy-candy, numeral-numeral and candy-numeral arrays within a single session, the chimps appeared to be unable to decipher the operative rules, and instead to adopt low-level response strategies such as a right-hand bias. Surprisingly, they did not show significant incentive interference effects in candy-candy trials or symbolic facilitation effects in numeral-numeral trials, but appeared to respond randomly. In mixed candy-numeral arrays, they tended to point to the numeral regardless of quantity. In effect, mixing array types within a single session seemed to confuse the chimps, suggesting limits to the extent to which numerals facilitate flexibly rational responses.

A final study was conducted in which different trial types were not mixed within a given test session, to see if the original interference and facilitation effects would again be found. They were: candy-candy sessions produced interference; numeral-numeral sessions produced facilitation. But mixed trial sessions revealed no consistent group pattern. More work is needed to clarify the conditions under which numerals facilitate rational responses by chimps.

(b) Reaction time effects of combining numerals with symbols for "same" and "different". In another experiment, Boysen combined numerals with symbols for "same" and "different", with which the chimps were also already competent from other work. They were asked to use the symbols for "same" and "different" to report whether quantities matched or not, and their reaction times were measured. The two quantities were presented as two arrays of dots, as two numerals, or as a combination of dot array and numeral. The aim was to probe and compare possible counting and subitizing processes. Subitizing is the rapid, accurate, perceptual apprehension of small quantities up a limit of between 3 and 5, as opposed to the slower, serial enumeration involved in counting larger quantities. The chimps could all use numerals "0" to "6", and some could extend to "8", so whether they used the faster or slower process might be expected to vary with quantities compared or with mode of presentation (dots vs. numerals) and to be revealed by reaction times. The animals were first trained on the task dot-dot, numeral-numeral, and dot-numeral pairs but only using the quantities 2 and 5. They were then trained similarly on 1 and 6, and then on 1, 2, 5, and 6. They were then tested on novel combinations using quantities 0, 3, and 4 as well. Their "same" judgements were faster for smaller arrays of dots than for larger arrays, but their "different" judgements were significantly faster than their "same" judgements for both dots and numerals, even for smaller quantity "same" comparisons. Reaction times tended to be longest for numeral-numeral comparisons and shortest for dot-dot comparisons, but these differences across stimulus types did not reach significance. Thus, the quickest reaction times are for judgements of different quantities, whether presented by means of dots or numerals, rather than for comparing smaller quantities or for comparing dots as opposed to numbers. No symbolic facilitation effect is shown, nor is it clear how the subitizing/counting distinction might explain this intriguing result.

(c) Solving problems with scale models. The third set of experiments reported by Boysen investigates chimps' abilities to find hidden objects by using scale models and other representational media, such as photographs and videos. In similar studies with children by De Loache, the child watches as a miniature toy is placed in a scale model of the playroom

and the child is then asked to find the real toy in the real playroom, which is hidden in the corresponding place. Three-year-olds were successful, though children only six months younger had difficulty; however, both age groups were successful when photographs were used instead. It was suggested that the scale model was harder for the younger children to interpret as a representation of the real world than the photos were, because the scale model had the additional aspect of a separate toy-like object, which the photographs lacked.

Boysen investigated chimps' ability to solve the hiding game using a scale model as well as other representations of the hiding place, including photographs and video. One of her chimps, Sheba, readily solved all variants of the problem, while another, Bobby, failed all variants and perseverated in various ineffective search strategies, such as circling the room clock-wise. When the hiding game was extended to include other chimps, a sex difference emerged: three females were successful even when the hiding sites in the model and its referent were rearranged between trials, while the males performed poorly and perseverated in the way Bobby had. Boysen suggests that these male individuals could not meet the attentional demands of representational processing, so could not control their impulsive, lower-level dispositions by using the representational strategy successfully employed by the females. She speculates that a metarepresentation bridging the scale model and its referent might provide a facilitating symbolic link that would enable the animals who failed the task to inhibit interfering responses, as in the candy-numeral studies, and plans further experiments to address this possibility.

The three lines of experiment together indicate that symbols facilitate rational responses and/or the inhibition of interfering dispositions in some situations, and for some individuals, but not others. While more work is needed to understand how and when symbolic facilitation operates, these results indicate the powerful effects symbols can have in modulating dispositions, and, as Boysen suggests, their adaptive role for early hominids.

1.7.4. Savage-Rumbaugh, Rumbaugh, and Fields. Primatologists Sue Savage-Rumbaugh, Duane Rumbaugh, and William Fields provide an overview of language research conducted with apes over four decades, aiming to delineate the circumstances that promote linguistic capacities. As animals acquire language, they can display evidence of rationality that is problematic to obtain otherwise. But learning language requires learning symb-ols that are freed from their context of acquisition to be used in novel ways, not merely learning associations.

The authors begin by providing a useful comparison of the methodologies of nine different animal language projects. The projects may use sign language, lexigrams on a touch panel, and/or spoken English; they may emphasize symbol production or comprehension; they may use traditional associative training techniques, alternative training techniques (such as the model-rival system), or no explicit training at all but immersion in social activities including language use. The authors argue that the differences among these projects in various methodological dimensions hinder direct comparisons of the skills that result, which also differ in various dimensions. The authors' research on ape language divides into four phases, employing different methodologies that are informed by experience in previous phases.

(a) Lana. In the first phase, the chimpanzee Lana was trained using geometric lexigrams embossed on a keyboard, which were displayed in the sequence touched above the keyboard. She was taught "stock sentences", strings of lexigrams following specific grammatical rules. Real progress began when her caregiver Tim entered her chamber and began to interact with her using the keyboard, and turned on processes of social learning rather than the rewarding of motor responses. For example, Lana was able to form novel sentences, recombining elements from five different stock sentences, to get Tim to share his coke with her when her vending machine was empty. No previously learned associations or behavioural invariances could explain this behaviour. In contrast with Herman's dolphins, who understand novel symbol combinations in grammatically appropriate ways, Lana was herself producing such novel sentences. Moreover, when first shown an object she didn't know the name of, and no one bothered to tell her, Lana spontaneously asked "what name of this?" These innovations indicate that Lana understood the limitations of her stock sentences. However, some of her sentences were not interpretable, and her comprehension skills were limited.

(b) Sherman and Austin. Phase two focussed on achieving communication between chimps rather than human/chimp interactions, and used single words rather than stock sentences, in order to try to understand how language might have evolved without continual assistance by competent language-users. Using the lexigrams keyboard, and emphasizing both production and comprehension, Sherman and Austin were taught to name items, to request items and to give specific food items to the experimenter in response to requests—which they found difficult initially, as they were strongly biased to attend to the foods they wanted to eat rather than to those symbolically requested. When they'd learned to do so, they were given the opportunity to request foods from one another. They initially failed to communicate, lacking the needed nonverbal coordination skills. But they began to succeed in verbal communication as they learned nonverbal gestural, timing, turn taking and attentional skills that coordinate conversation. For example, if Austin were looking away when Sherman selected a symbol, Sherman would wait until Austin looked back, point to the symbol, point to the object it referred to, and so on. They appeared to recognize that speakers must monitor listeners' attention and comprehension, and repair conversations as needed.

In another test, one chimp's keyboard was inactivated and he was provided instead with brand labels from food packaging by means of which he could communicate the contents of a container to the second chimp, who could then request it using his keyboard. Both chimps did this from the first trial. They were not trained or shown how to use the brand labels as symbols for food or to associate them with lexigrams, but were merely provided with the materials. No previous experiences explain this innovative use of labels as symbols; rather, the chimps used them to communicate information they recognized the other lacked, indicating understanding of the mental state of the other.

Moreover, Sherman and Austin appeared to recognize that differential mental states existed more generally; they spontaneously began to announce what they were about to do before doing it and to report unusual events selectively to others who had not witnessed them. They also engaged in pretence games, showing what Currie would call enrichment, for example, by pretending dolls had bitten their fingers and then nursing their fingers. As their television-watching skills developed, they began to differentiate between live and taped images of

themselves, watching the live images of themselves monkeying about and using the live images to investigate their own bodies. The authors suggest that these behaviours display an increased capacity for self-reflection as well as for understanding the minds of others.

(c) Kanzi. Phase three focussed on the linguistic capacities of bonobos rather than chimpanzees, beginning with an adult female Matata. Her training was the same as that Sherman and Austin had received, but her progress was extremely slow. However, her son Kanzi was allowed to remain with her during training sessions, from age 6 months to 2 years, at which point he was separated from his mother temporarily. Suddenly, his latent linguistic learning was activated; he used the keyboard over 300 times in the first day of separation. In the first month of separation, Kanzi stated his intentions, shared food, combined symbols spontaneously, and replied to requests symbolically. By age three he had acquired all of Sherman and Austin's trained skills, though he had never received training himself but only observed attempts to train his mother. Rather than receiving training, Kanzi was immersed in language in ordinary life; he acquired new lexigrams rapidly along with recognition of corresponding spoken words.

By age 8, Kanzi could interpret novel English utterances that were syntactically complex and semantically unusual, such as "Get the toy gorilla and slap him with the can opener" or "Feed the doggie some pine needles". Formal tests of this ability indicated that Kanzi had broken the syntax barrier and was simultaneously processing both syntax and semantics. For example, when confronted with a collection of objects including a knife, a mask, an apple, a toy dog, a balloon, a cup of water, a TV, juice in a bowl, and an egg in a bowl, Kanzi responded correctly both when asked to pour the egg into the juice and when asked to pour the juice into the egg. He responded correctly to novel recursive requests such as "get the tomato that is in the microwave", ignoring both the tomato in front of him and the other items in the microwave in doing so, and to questions such as "can you feed your ball some tomato?", finding a previously ignored face on his ball and placing the tomato on its mouth. The authors argue that such responses cannot be explained as responses to single words in the order presented or in certain contexts, without an understanding of syntax.

(d) Panbanisha and Panzee. Was the ease with which Kanzi acquired language compared with Lana, Sherman, and Austin due to immersion in language use, or to his species? In phase four, this question was addressed by co-rearing a chimpanzee, Panzee, and a bonobo, Panbanisha, from infancy to four years in a language immer-sion environment. Both began to "babble" at their keyboards at about 6 months of age, but Panbanisha began using lexigrams to communicate at an earlier age than did Panzee. By age 4 years, the bonobo understood 179 words and lexigrams, including many proper names for chimps and people, while the chimp understood only 79, including no proper names. Moreover, the chimp often over-generalized her symbols, failing to mark distinctions that the bonobo did mark. The chimp's comprehension was more closely linked to the here and now and presently displayed objects, while the bonobo responded appropriately to remarks about the future or the past, and was relatively more concerned to communicate about social relationships than about tools.

The authors summarize the most important findings of their research as follows. First, language can be readily acquired by apes under three years of age, without any training, by

immersion in a social environment where English is spoken normally to provide a narrative of everyday life and of social interactions, and where lexigrams accompany spoken words. Second, lexical production emerges spontaneously from untrained comprehension, but comprehension does not emerge spontaneously from the trained production of words in the right context. The latter tends instead to produce learned associations bound to a specific context. Such training may actually inhibit acquisition of a fully representational language that can be used freely across contexts and to communicate about what is absent or past. Third, when an ape achieves a level of symbolic understanding that enables it to employ symbols freely to state its intentions and to report events, many other rational capacities begin to be displayed as well, including capacities for directing attention by pointing, for pretence, for self-reflection, and for mind reading.

In the authors' view, acquiring language brings with it the understanding that it functions to communicate information, and hence the realization that the contents of others' minds can differ from those of one's own mind. The authors share Povinelli's view that current nonverbal assessments of animal mind reading are inconclusive. But they argue that linguistically competent animals can instead be asked what they are thinking about, under conditions in which the questions and answers are novel and could not be learned responses. The bonobos reared in language immersion environments answer this question appropriately in various contexts and validate their reported thoughts with appropriate behaviour. The authors report, moreover, that Kanzi and Panbanisha pass standard verbal false belief tests normally used with children. Language acquis-ition thus makes it possible to assess aspects of animal rationality, evidence for which would otherwise be problematic.

Finally, the authors endorse the theory of Greenspan and Shanker that symbolic communication emerges from emotional signals in interactions between infant and caregiver. They suggest that the paradigms likely to produce the highest levels of linguistic competence in animals are those that foster the closest relationships with their caregivers, a prediction confirmed by ape language results.

1.8. Why does it matter?

Why, then, does it matter whether animals are rational? It matters both for our understanding of other animals and of ourselves. We live with animals, we interact with them and use them in our daily lives, and share a planet with them. Yet we see ourselves as discontinuous from them in important ways. Rationality is one of the main hooks on which human distinctiveness and specialness has been hung. We treat rationality as having intrinsic worth, in addition to sentience. If a creature can feel pain, we feel we ought to avoid making it suffer unnecessarily, but we may not on that account grant it the additional intrinsic value and dignity associated with rationality. Understanding whether and in what ways nonhuman animals can be rational may prompt us to rethink human rationality, our relations to other animals, and our own irrationalities. Perhaps our rationality is more piecemeal, less theoretical, more embedded in and conditioned by our environments than we realize. The possibility of explaining relatively complex animal behaviour by appealing to a variety of relatively simple, domain-specific processes may lead us to re-evaluate our presuppositions about human rationality and to ask whether we are operating a double standard in assessing

human vs. animal rationality (see and cf. Shettleworth 1998, 563). On a disaggregated view of rationality, there is no single boundary distinguishing the human mind from other animal minds in respect of rationality; rather, there are various specific dimensions of comparison. Making comparisons across species and with human beings in this way elucidates the extent to which the abilities of different animals and of humans can be explained by appealing to similar kinds of processes, and in what sense these processes are rational. Should this rethinking debunk human rationality? Not necessarily. Rather, it may help us to understand it better, and to understand the continuities as well as the discontinuities between human and other animals.