



UNIVERSITY OF LEEDS

This is a repository copy of *Animal Agency*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/84728/>

Version: Accepted Version

Article:

Steward, H orcid.org/0000-0003-1654-577X (2009) *Animal Agency*. Inquiry: an interdisciplinary journal of philosophy, 52 (3). pp. 217-231. ISSN 0020-174X

<https://doi.org/10.1080/00201740902917119>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Animal Agency

HELEN STEWARD

University of Leeds, UK

ABSTRACT Are animals agents? This question demands a prior answer to the question what an agent is. The paper argues that we ought not to think of this as merely a matter of choosing from a range of alternative definitional stipulations. Evidence from developmental psychology is offered in support of the view that a basic concept of agency is a very early natural acquisition, which is established prior to the development of any full-blown propositional attitude concepts. Then it is argued that whatever one makes of the developmental evidence, it is in any case arguable on other grounds that the concept of agency as we have it in adulthood remains perfectly comprehensible independently of any reference to the more sophisticated propositional attitudes. Any reluctance we might feel to ascribe such things as beliefs to non-human animals, therefore, need not stand in the way of the claim that they are agents, nevertheless. The paper attempts to characterise the core of this basic agency concept, and discusses, albeit briefly, the question how we ought to decide which animals are to be thought of as falling under it. It concludes with some speculations about the nature of the intellectual currents which have made the shape of this important concept so hard for us to discern.

Correspondence Address: Dr. Helen Steward, Dept. of Philosophy, University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, UK. E-mail: h.steward@leeds.ac.uk

The question whether any non-human animals are agents might look at first sight, to a range of different kinds of philosophical eye, like a rather uninteresting one. Some, I think, might suppose it uninteresting because they imagine it to be easy – that the answer is simply a very obvious ‘yes’. Animals do such things as build nests and burrows, seek food, attempt to elude predators. The more sophisticated among them appear to communicate with one another, even play. They fight. They groom. If the squirrels in my garden are anything to go by, they have a range of sophisticated problem-solving strategies available for confounding the best attempts of human beings to outwit them. Surely all these are actions, and so surely animals are agents? Does more need to be said? The more cautiously-minded, on the other hand, might suppose the question uninteresting in a different way. We are used in philosophy to questions which cannot be sensibly answered until various of the concepts contained in them have been defined – but once the various different possible precisifications are in, it is easy enough to see how the question should be answered on each of the different interpretations. And one might at first sight think the question whether animals are agents was one of these sorts of questions. For example, one might suppose that we could define an agent to be a creature possessing powers and capacities a and b – in which case some non-human animals might clearly qualify; or we might choose to require in addition that an agent must have capacities c, d or even e; in which case it might be evident that no non-human agent could possibly make the grade.

In fact, though, the philosophical controversy about animal agency does not conform to either of these pictures; indeed, the fact that there is much of a controversy at all would seem in tension with the presuppositions of the first. Certainly, it is not obvious to everyone that there are non-human animal agents – there are those such as

Davidson, for example, who deny, at any rate, that any creature without a language could be an intentional agent, and it is not clear whether he thinks there is any other kind of agency that would be relevant in the context of this debate.¹ But neither does it seem as though Davidson is merely utilising from the start a more demanding and sophisticated notion of agency than are those who think it preposterous to deny that animals are agents. It is not as though he simply explicitly builds into his definition of agency extra powers, which it is unreasonable to suppose any non-human animal to possess, powers which others would choose to leave out of their explicit definitions. His insistence is rather that a conception of intentional agency of which he takes it for granted we have a shared, basic understanding, is so connected to other powers and capacities – in particular to the capacity to have beliefs, and the capacity to ascribe beliefs to others – that we cannot coherently imagine how the basic power could be exemplified without the others. The way the debate has proceeded, then, makes it look as though the notion of intentional agency is assumed to be one on which we have a basic pre-theoretical grip, such that our disagreements about its applicability in the animal case are real, and not merely verbal ones.

I agree with this assumption. I think there is every reason for thinking that there is a robust and highly distinctive agency concept which deserves to be recognised as a hugely powerful organiser of human thought – and I shall say something shortly about the support that can be provided to this claim from developmental psychology. I also agree with Davidson that this concept of an agent is part of the ‘matched set’ which constitutes propositional attitude psychology – that it is a precondition, indeed, of any non-metaphorical use of concepts like belief, desire and intention, that an agent possessing those beliefs, desires and intentions, be presupposed. But it is much less clear, I think, that all of the full-blown propositional

attitude concepts must come along for the ride whenever the concept of agency is deployed, and in particular, I think it is not clear that a fully-fledged concept of belief must do so. I shall try, in what follows to paint a picture of what the concept of an agent is the concept of, and to support it with some evidence from developmental psychology, which does not make this presupposition. If it is right, there may be room for us both to accept what seems to me to be the extremely natural view that large numbers of animals are agents, and yet to concede, along with Davidson, that they may perhaps not have any states which it would be truly appropriate to call 'beliefs'.

The structure of the paper will be as follows. First I shall describe some ideas which have been put forward by developmental psychologists, seeking to understand the emergence in infancy of our theory of mind, which seem to me to be potentially highly relevant to the question what the structure is of our concept of agency. Several psychologists have argued that developmentally speaking, at any rate, the concept of agency is a more basic possession than some of the propositional attitude concepts proper, and I shall try to describe, in rough outline, some of the more prominent of these theories. Then I shall try to explain why, whatever one makes of the developmental evidence, it seems to me in any case to be arguable, for other reasons, that the concept of agency is comprehensible quite independently of any reference to the more sophisticated propositional attitudes and can be regarded, therefore, as unproblematically applicable to many non-human animals. I shall try to characterise the essential core of the concept of an agent, as I understand it, and I shall say something, albeit brief, about how we should go about deciding which animals might be thought of as falling under it. Then, finally, I shall try to explain why I think the shape of this basic concept has become surprisingly hard to discern, in our times, buried as it has been underneath the sediments deposited unhelpfully upon it by quite

a varied range of recent and not so recent intellectual currents. I shall also try to give an indication of why I think it is time we unearthed it from beneath them.

I. The Evidence from Developmental Psychology

The capacity to attribute fully-fledged propositional attitudes to agents is an impressive and complex ability which most psychologists believe does not begin to emerge until at least halfway through the second year of life, and which takes several years to come to full fruition. The capacity to distinguish animates from inanimates, by contrast, develops much earlier; as does what seems likely to be a related sensitivity to goal-directed actions. There is very strong evidence that from early infancy, human beings are disposed to process information about the movements of animate entities quite differently from the movements of inanimate ones, and that they bring to bear on these movements explanatory principles never normally looked for in the explanation of purely mechanical interactions.² Infants as young as seven months appear to make distinctions between things which can move themselves and things which require an external source of energy, such as a push or a pull from some other object, to make them move.³ Spelke, Phillips and Woodward (1995), for example, report an experiment which appears to show that seven-month-old infants do not apply the contact principle – that is to say, the idea that objects act upon each other if and only if they touch – to human beings, though there is a wealth of evidence to suggest that they reason in accordance with the principle where inanimate objects are concerned.⁴ Infants' interest (as measured by looking times) in a videotaped display in which one large, brightly coloured object with a 'meaningless' shape moved towards a second, stopping a little way short of the second, to be followed by movement in the

second object, was reliably greater than in an event in which contact occurred, suggesting that infants are surprised, in this case, by the violation of the non-contact principle. But no such effect was remarked when the same patterns of motion were presented, but where the objects in question were human beings. Here there seems to be no expectation that the objects will conform to the contact principle. Even at this early age, then, it appears that different causal principles are applied to the motions of inanimate objects, on the one hand, and to human agents, on the other.

Other evidence suggests, though, that it is not only human agents whose motions are differently coded and conceptualised by the developing infant. Gelman (1990) proposes that an early general distinction is made between animates and inanimates which is shaped and facilitated by what she calls two important ‘skeletal causal principles’, which she dubs the innards principle and the external agent principle, respectively. Processing governed by the innards principle, she argues, treats certain natural objects – the ones we call ‘animate’ – as though they have something ‘inside’ them which explains their movement and change. Processing governed by the external agent principle, on the other hand, is constrained by the expectation that movement in a given object requires the imparting of some sort of external impetus or force.⁵ These skeletal principles, Gelman proposes, work together with complementary perceptual processes which focus attention on features either innately known, or swiftly learned to be relevant to the categorisation of objects as animate or inanimate (such as presence of faces, biomechanical patterns of movement, surface cues indicating the nature of the ‘stuff’ of which a thing is made), in order to aid in the rapid accumulation of knowledge about the animate-inanimate distinction.

It appears, though, that the presence of these extra perceptual features is by no means required in order to cue the type of processing associated with the discernment

of goal-directed action.⁶ Research carried out by Gergely et al. (1995) appears to demonstrate that infants as young as twelve months old interpret computer animations in which a small circle approached and contacted a large circle by jumping over an intervening obstacle, in a way which reveals that they apply what Gergely and Csibra (2003) have called ‘the teleological stance’ in such cases. The teleological stance, as characterised by Gergely and Csibra, is non-mentalistic in that it does not involve the attribution of true intentional states to the agents involved. Rather, the stance is based on relationships between three crucial representational elements: the action itself, a possible future state (the goal) and the relevant situational constraints. Given information about two of these three elements, infants are able to make an inference about the third, based on what Gergely and Csibra call ‘the rationality principle’, which supposes that agents will in general take the most efficient action for achieving their goals, given the situational constraints as these are perceived by the infant herself. For example, in one of the violation-of-expectation studies conducted by Gergely et al, infants who had been habituated to the computer animation described above were then tested to see what happened if the ‘situational constraint’ (the intervening obstacle) was removed. Infants were shown two test displays: the same jumping goal-approach as they had seen during the habituation phase (only now without the obstacle), or a novel straight-line approach. Infants looked longer (indicating violation of expectation) at the old jumping action – arguably because they regard this as an inefficient means to the goal, in the absence of the obstacle. But crucially, no attribution of intentional states to others is required in order to apply the teleological stance. All that is required is the simple postulation of an action with a goal. It is thus a simpler stance than the intentional stance – and does not require any understanding of such things as pretence or false belief.

It seems very plausible that the basic distinction between things which move themselves (as the objects in the animated display at least appeared to), and things which require some kind of external impetus in order to move, provides the initial criterion of selection (or at any rate, one of the most important criteria) by means of which objects are or are not treated as potential candidates for inputting into the further module or modules which embody the infant's nascent theory of mind. This view is found in the work both of Premack and of Leslie, though their accounts of the structure of the mind module differ in various respects.⁷ According to Premack's theory, the infant divides the world into two kinds of object, those that are, and those that are not self-propelled, and this then becomes the basis of further differentiated interpretation of the movements in question:

... first, the infant perceives certain properties, for example, one object is moved by the other under conditions of temporal and spatial contiguity, versus the object is self-propelled ... second, the infant not only perceives, but also interprets, that is, the infant's perception is the input to a slightly higher-order device that has interpretation as its output. The interpretations in the two cases in question are causality and intention respectively.⁸

Leslie's early view is somewhat similar, in that he, too, supposes that infants are innately predisposed to pay particular attention to the question whether an object is or is not made to move by something else, and that this distinction then serves as a means of selecting inputs to a higher level subsystem which he calls ToMM (Theory of Mind Mechanism). For Leslie, though, there are reasons for dividing ToMM into two parts: one concerned with agents and the goal-directed actions they produce, the

other, more sophisticated module with propositional attitude ascriptions. What Leslie calls ToMM system₁ begins to develop around six to eight months – one of the first signs of its development being the following of eye gaze. System₁ can represent an agent as acting to bring about a state of affairs and thus recognises the basic notion of a creature's having a goal or end. Other indications of the emergence of system₁ include the fact that infants begin, during the second half of the first year, to acquiesce and help in achieving the recognised goals of others – e.g. in positioning to be picked up or to have a nappy changed – and they also begin requesting behaviours – e.g. handing an object to its mother as a request to operate it. But Leslie argues that this concept of goal-directedness is not yet a propositional attitude notion, because 'acting to bring about ...' does not describe an attitude to a proposition, but rather an endeavour to change physical circumstances.⁹ Propositional attitudes proper are the preserve of what he calls ToMM system₂ which does not begin to develop until during the second year of life, and involves the attribution of full-blown representational states, such as beliefs, to others. The most obvious early sign of the emergence of this system, according to Leslie, is the ability to pretend, and also to understand pretence in others, which normally emerges between eighteen and twenty-four months. Eventually, ToMM system₂ presumably develops into what underwrites the normal capacity to apply the intentional stance – a perspective from which behaviour is explained and predicted by means of the attribution of such things as beliefs, desires and intentions to agents.

Suppose some story of this kind were correct in broad outline. If it were, it would follow that non-self-movers, on the whole, would not be considered to be contenders for possession of mind at all.¹⁰ Self movers, though, would go forward, to be inputted experimentally into the developing mind module(s), which then attempts

to assign a goal to the candidate agent. Not every self-mover may be able plausibly to sustain its candidacy for very long, of course. Gelman, Durgin and Kaufman note that machines represent an interesting hybrid category, since though they appear to move on their own, they are made of inanimate material, do not exhibit typical biomechanical motions and do not adjust well to local environmental problems:

Robots are not particularly good at adjusting their motions to local perturbations in the environment; in contrast, so predictable is the animate world's ability to deal with unanticipated holes, bodies of water, oil slicks, branches that come below the head, sun in the eyes, weather changes, etc., that we almost forget how remarkable are the action abilities of the animate world. This reflects the fact that machines do not exhibit the kind of action that is controlled by biological mechanisms ... action patterns that are noted by infants ... and used as cues for animacy. Additionally, machines are made of the wrong stuff.¹¹

These authors are admittedly primarily interested in the child's conception of animacy, not agency – but it is not implausible that the problems robots generally swiftly encounter in dealing with obstacles renders them unsuitable candidates for genuine agency, too. Initially-formed hypotheses about purposiveness, for instance, might be hard to maintain in the face of a robot's very limited abilities to keep itself on course despite deflections and obstacles – one cannot readily tell what might be the aims of something which cannot devise ways to overcome quite simple problems in order to achieve them. ToMM system₁, it is reasonable to suppose, can only really get going properly where the objects with which it is confronted are good at attaining

their goals – and robots, even when such ‘goals’ are programmed in, are not terribly good at it. Added to the fact that robots do not move in the ways typical of the animal agents in response to which ToMM must originally have evolved, and are generally made of a very different kind of material, it would perhaps not be surprising if, as Gelman, Durgin and Kaufman speculate, they are eventually assigned to a category of their own – self-movers, perhaps, but not necessarily ones to which coherent goals can be readily or profitably ascribed, and almost certainly not ones which deserve the further attentions of ToMM system₂. This initial prejudice, presumably, might later come to be confirmed when, as older, more mature operators of the ‘innards principle’, we come to recognise that although there is indeed something inside a robot which guides and controls its behaviour, that something is a very different sort of something from what is inside an animal – a ‘something’ which, we are inclined to suppose, requires no subjective perspective, no true intentional states and certainly nothing amounting to spontaneous injections into the course of nature on the part of a robot-self.

This sort of account fits well with more recent proposals by Biro and Leslie (2007), who offer what they call a ‘cue-based bootstrapping model’ of infant development, according to which an initial sensitivity to certain behavioural cues for goal-directed action leads to learning about further cues, which in turn feed back to force revisions, modifications and developments of the original cue-based system. On this view, the infant begins with a core notion of a goal-directed agent that is triggered by a certain range of cues (for example, self-propelled motion, equifinality¹² and an action-effect). Infants then gradually begin to calculate the statistical associations between these cues, and between these and other properties of the events they witness – such as, for example, the general appearances of things likely to exhibit the cues.

Once such associations have been learned, infants can anticipate goal-directed actions without direct benefit of the cues and can also reject certain candidacies, despite the superficial evidence provided by the cues. In this way, we might conjecture, children become able, eventually, to discern an action even in the absence of any type of motion at all (e.g. when someone is deliberately keeping still for the purposes of a game of musical statues); and also to decide that what they have witnessed is not one, despite the presence of the cues (as e.g. with a computer animation in which one circle ‘pursues’ another). The cues are thus not criteria for the application of the developed concept – they are rather clues utilised by the mind module en route to the establishment of a more mature conceptual capacity.

Developmentally, then, there seems to be a great deal of evidence supporting the view that a basic conception of purposive agency is in place prior to the emergence of full-scale propositional attitude psychology. Of course, it might be suggested that the basic conception is simply replaced in the developing child by a richer notion, once propositional attitude psychology develops. But actually, I think there are reasons for thinking that that is implausible – and that the more basic conception survives into adulthood, where it continues to structure our thinking and our perceptions. Propositional attitude psychology, I want to argue, is not the essence, but rather a sophisticated outgrowth of the basic concept of agency, suited in particular to enable us to deal with our human conspecifics. The concept of an agent is a more general and less demanding notion, applicable unproblematically to many sorts of animal, as I shall now go on to argue.

II. The Evidence from Everyday Thinking

Consider the following quotation from Wittgenstein's *Philosophical Investigations*:

We say: 'The cock calls the hens by crowing' ... Isn't the aspect quite altered if we imagine the crowing to set the hens in motion by some kind of physical causation? (1953, §493).

Wittgenstein here draws attention to the way in which the way in which our perceptions of animal activity are drenched in presuppositions of agency. The aspect is altered if we imagine the crowing to set the hens in motion by some kind of physical causation. But why is that? What is it that we imagine happens between cock and hens in the first place; what is it which has to go once the aspect is altered by the new suggestion? It is too much, surely, to suggest that we suppose that the cock wants the hens to come over and believes that by crowing he will be able to encourage them to approach; and that the hens, hearing him, understand that he is thus encouraging them and want to do as he asks. That would be what a truly full-scale importation of propositional attitude psychology would involve, and surely we need not be thinking anything quite so laden with ideas taken from the paradigm of linguistic communication as that, before we are able to feel the power of Wittgenstein's question. But there are, nevertheless, a number of things present, I suggest, in our thinking about animals and their activities which, though not amounting to the attribution to them of full-scale propositional attitude psychology, are nevertheless peculiar to the animate world, and which must therefore be made to disappear once their interactions are supposed reduced, by an imaginative aspectual shift, to purely physical causation.

One thing that we immediately impose, for example, on our conception of an animate entity, at least at first, is a dualistic scheme of animal and body, according to which we suppose the animal to be a possessor of its body, in much the way that we are possessors of our own. There is the cockerel, and there is the cockerel's body which the cockerel can make move, just as we can make our bodies move. 'He is puffing out his chest', we say, or 'he is fanning his tail'. No such distinction is ever required or imposed, I suggest, in connection with anything inanimate; it is a privilege reserved, except where we are speaking merely metaphorically, for the animal kingdom. It is connected, doubtless, with the presupposition that an animal is a centre of some form of subjectivity, subjectivity which affects and mediates its interactions with its environment. And we are, moreover, I think, prepared to ascribe a range of psychological states to a cockerel, even though they are not perhaps the ones I mentioned above. If the cockerel moves suddenly from one side of the farmyard to another, for example, we might hypothesise that that was because it had seen a dog through the gate, or because it wanted shade, or because it was trying to get nearer to the hens, which had wandered off in that direction. But none of these explanations invokes a real propositional attitude. Indeed, I am not sure the states we most naturally turn to in order to describe the mentality of animals are really propositional attitudes at all. They are such things as seeing, wanting, and trying to get, none of which is associated with a verb which needs to take a proposition as complement, and the latter two of which positively have to be forced into the propositional attitude mould – to make verbs like 'want' and 'try' take a 'that' clause requires some effort.

Given that all this is in place, then, there is obviously some reorientation involved in accommodating the thought that in fact it is purely physical causation which sets the cock crowing and then in turn sets the hens in motion, or which makes

the cockerel set off across the farmyard. It is most unnatural to suppose that the cockerel was caused to make its journey across the yard by anything like a mere reflex or a simple stimulus-response mechanism. For although we obviously have to recognise the huge importance of instinct in the lives of animals, instincts which prescribe for a given animal a range of basic activities from which it is certainly not free to forbear, I think we allow to the animal – and this is crucial, in my view, for the concept of agency – a certain freedom and control over the precise movements by means of which it satisfies those instinctual needs and desires. It decides, we think, precisely where it will go in search of food or shelter or to evade predators. Our natural inclination is to think of an animal as a creature that can, within limits, direct its own activities and which has certain choices about the details of those activities. To invoke a terminology I find helpful in this context, it is natural to think of animals as the settlers of various matters which concern the movement through time and space of their own bodies, and I submit that it goes deeply against the grain to suppose that each exact detail of each movement orchestrated by an animal was settled at any point prior to a period broadly coeval with what we think of as the period of the animal's action.

Putting these various things together, we might then say the following about the concept of agency which is at work in our understanding of animals:

- (i) an agent can move the whole, or at least some parts, of something we are inclined to think of as its body;
- (ii) an agent is a centre of some form of subjectivity;
- (iii) an agent is something to which at least some rudimentary types of intentional state (e.g. trying, wanting, perceiving) may be properly attributed;

(iv) An agent is a settler of matters concerning certain of the movements of its own body i.e., the actions by means of which those movements are effected are considered to be non-necessitated events, attributed always first and foremost to the agent, and only secondarily to environmental impacts or triggers of any sort.

To say that the concept of agency is in fact applied by us in connection only with animate entities is emphatically not to say, that nothing made of inanimate matter could ever conceivably constitute an agent. Being made of the right kind of stuff, and moving in characteristically biomechanical ways are doubtless important cues for the attempted application of the agent-concept, but it does not seem likely that they are criteria for the application of the developed concept. The test for this can only really be what people are inclined to think when asked – and my own experience with generations of students suggests that on the whole, people are inclined to agree that what a thing is made of and what it looks like can in the end not be allowed to count definitively for or against its agency or mindedness, conceptually speaking. Most agree that there is no a priori reason why an artificial system which could display sufficient flexibility and intelligence in its responses to its environment might not eventually demand interpretation as a conscious wielder of agency. There may in fact be empirical reasons, of course, why only certain sorts of matter can support the phenomena associated with consciousness, subjectivity and the capacity to settle things; indeed, it seems to me quite likely, as John Searle has argued, that consciousness will turn out to be essentially a biological phenomenon, as will agency, in my robust sense of that term. But we are not yet in a position to be able to be sure that this is so. And there is nothing in the concept of agency itself, I believe, which

rules out the possibility of its instantiation in artificially created, non-biological matter. If this is an impossibility, it is an empirical, not a conceptual one.

Neither is it the case that the agency concept is applicable to all animals, despite its deep connections with the animate/inanimate distinction. It does not follow from the fact that the agency concept develops out of information processing systems designed to conceptualise the movements of animal entities in a special way, that we cannot, as more mature wielders of the concept thus formed, then ask the normative question whether the concept truly applies to creatures of a given sort. The concept, once formed, has its own integrity; for it is indeed a concept that has been formed, and not a mere disposition. It is perfectly possible, therefore, after reflection, or empirical investigation, to decide that a given type of animal does not meet the conceptual criteria for agenthood, because it has been discovered in the case of that animal that what perhaps to a first view looked a bit like a case of agency, in fact does not qualify. A paramecium, for instance, might move in such a way as initially to suggest the applicability of the agent-concept and to cue its application in young children (or even in adults) watching the movements through a microscope, say (just as the kinds of animated computer displays described above cue the attempted application of the scheme, in ways which leave a distinctive phenomenological trace when one perceives such displays, despite the fact that as an adult, one knows perfectly well that the little circle is not really trying to contact the larger one). But once we find out that a few simple equations govern the movement of the paramecium through the water, there is no reason to suppose that there is any role left for a paramecium-self to play in the control of the paramecium body. There is, in fact, only the body, actuated by a range of forces. And this means that the paramecium cannot be an agent, according to the agent-concept described by (i)-(iv). None of its interactions with the environment

need be mediated by anything desire-like, and none need involve anything like an act on the part of the paramecium. Everything can be arranged by means of simple chemical processes, and the postulation of mentality under these circumstances is surely *de trop*.

What the paramecium case shows, though, is the vulnerability of the untutored perceptually-based intuitions which select potential inputs for the mind module, to the results of subsequent reflection and enquiry. The paramecium looked as though it might have had a mind; but it turned out that it did not. And having made this concession, of course, it might seem as though we are at the top of a slippery slope. Might we not come to have reasons to think we need to revise untutored views also of animals rather more complex than the paramecium? Might not such things as worms and insects, say, succumb to a similar treatment? Indeed, why stop there? Perhaps even the agency of cows and sheep will turn out to be comprehensible without any essential invocation of thought or mentality of any sort. And what about our own agency? We too are actuated by chemical processes and neural mechanisms. Once we have given up on the paramecium mind, is there anything to stop the whole of folk psychology collapsing like a giant house of cards, as eliminativists have always suggested it will, in the face of the advancing neurobiological sciences?

I think the answer to this is ‘yes’ – there is plenty to stop it collapsing. There are many reasons for supposing that the agency scheme is essential to our understanding of the animal world and I find it curious that it is so often assumed, even when this point about its essentiality in practice is conceded, that its essentiality in principle must be quite another matter. For why should one suppose this? Is it not natural to suppose that as we proceed upwards from the paramecium to ascend the ladder of biological complexity, that we encounter systems whose behaviour is

increasingly dominated by the genuinely special mode of functioning we call action? – and which must therefore be understood in terms of the agency scheme, because agents is what they are? This view is surprisingly unpopular. But why is it unpopular? What prevents people from being readily able to accept that the scheme is anything more, particularly when it comes to non-human animals, than a useful heuristic device?

I think an immensely powerful alliance of intellectual forces have conspired against the view that animals could truly be agents. Some are forces of reaction – the many cultural and religious forms of anthropocentrism which have long encouraged us to make as many clear separations as we are able between ourselves and our animal relatives. But many – perhaps more, in the end - are actually the (in my view) misdirected forces of scientific progress, enlightenment and modernity. One culprit, for example, is the confusion on the part of many naturalistically inclined scientists and philosophers of the essential and important dualistic scheme of agent and body which I have tried to describe, according to which the agent is conceived of as a thoroughly embodied entity, with the Cartesian dualist scheme of mind and body, which is rightly seen as thoroughly disreputable and to be avoided at all costs, but which talk of ‘agents’ is sometimes confusedly thought to presuppose. Another is the idea that where lower levels of explanation are available they supplant the higher – that explanations of animal movements are available at the levels appropriate to neurology, microbiology and perhaps even physics, that these can and must replace the agency scheme. Yet a third problem is an empiricist epistemology which has encouraged throughout the biological sciences a huge distrust of the spontaneous codings of the mind module and has supported the insistence that they be treated as nothing more than misplaced anthropocentrisms. And Kantianism has not helped,

either – for it has made prominent conceptions of agency in which a very rich conception of rationality has been privileged over more basic ideas like those of body possession and bodily control, ideas which in my view form the true heart of the concept of agency.

It is not part of my purpose in this paper to take on these forces. And so for all I have said, perhaps it remains a possibility that no non-human animal could truly be an agent in the sense implied by (i) – (iv) above; perhaps indeed that not even any human animal could be one. These are issues I am not able here to address. But my concern is not so much to show that animals are agents – as to insist that some part of us finds it almost impossible not to categorise them as such, and to plead that this intuition deserves not to be lost underneath the mountain of epistemological scrupulousness, mechanomorphism¹³, anti-dualist fervour, and behaviouristic scientific methodology that has, I believe, obscured our view of the concept of agency. It seems to me that the results of the primitive and unreflective categorisation of animals as agents are phenomenologically tangible to those open to its influence – watching a bird pecking around for food or a cat stalking a mouse is just utterly unlike watching, say, trees blow in the wind or a car drive down a road. To watch a creature engaged in such goal-directed activity is, I maintain, to think of it as a moment-to-moment controller of its own body, a centre of subjectivity, a possessor of some representational and some motivational states (whether or not we are prepared to call these ‘beliefs’ and ‘desires’) and a settler of matters which concern its own bodily movements – and this way of thinking is at the same time a way of seeing. These intuitions undoubtedly become less powerful as the animals in question become less easy to categorise as animate to begin with (e.g. where facial features are absent or difficult to perceive) or where many of a given creature’s movements cannot easily be

interpreted as the results of intentional, goal directed actions (e.g. where it is obvious that some ‘tropism’ or other is really responsible for some movement or other, as when a moth flies repeatedly into a light); or where it is just more difficult, owing to factors to do with size, speed and habitat, for us to watch a creature’s activity over the period of time that might be necessary to reveal the goal-oriented patterns in its behaviour. And the concern that perhaps the truth is that the strength of the intuitions diminishes as a creature becomes less like a human being, and that this might not really constitute an adequate basis on which to make or deny ascriptions of mindedness is, of course, a legitimate one. But before we embark on the task of interrogating our intuitions, we should at first allow them to register with us. For perhaps it may be that they make distinctions in the places they do for reasons not altogether unconnected with the fact that there are indeed distinctions in those places which need to be discerned.¹⁴

¹ Davidson (1982).

² See e.g. Premack (1990), Gelman (1990), Leslie (1994), Gelman, Durgin and Kaufman (1995), Spelke, Phillips and Woodward (1995).

³ See Premack (1990), Leslie (1994).

⁴ See Leslie (1982, 1984), Leslie and Keeble (1987), Van der Walle and Spelke (1993).

⁵ See also Leslie’s account of the module he called ‘ToBY’ (Theory of Body), which amongst other things, assigns a source of energy to any given motion according to the answer to the question whether it was made to move by something else, or alternatively, has an internal source of energy.

⁶ Though the matter remains controversial – see Woodward (1998) for a view opposed to that outlined in what follows. It should be noted, though, that this controversy is orthogonal to the main concerns of the present paper; for both camps agree that the development of the agency concept is early, and precedes by many months the establishment of any facility with propositional attitude attribution.

⁷ Premack (1990), Leslie (1995).

⁸ Premack (1990), p. 3.

⁹ For similar views, see Golinkoff (1983), Poulin-Dubois and Schultz (1988).

¹⁰ What about dolls, teddy bears, and the like? While the tendency of young children to attribute mental states to their toys may be testament to the power of facial features to serve as a preliminary guide in the sorting of objects into those which are and those which are not potential candidates for application of theory of mind (and see Johnson (2000) for further evidence), it seems fairly clear that children are aware from a pretty early age that dolls, and the like, are only ‘pretend’ subjects. In a study by Massey and Gelman (1988), for instance, children were clear that figurines, statues, and the like could not move themselves up or down the hill, and justified this claim, when asked, by appeal to the fact that these things were (i) not real, or (ii) made of the wrong ‘stuff’. Though faces matter, then, and though Leslie is surely right when he speculates that “it seems highly likely that such low-level mechanisms have inputs into ToMM” (1995, p.145-6), it seems likely also that ToMM quickly learns to override the prima facie evidence of mindedness that is supplied by a face. Children swiftly learn, in other words, that a face is not a sure sign of a self-moving animal, and nor, therefore, a sure sign of a mind.

¹¹ Gelman, Durgin and Kaufman (1995).

¹² ‘Equifinality’ is a property possessed by an act-type when it may be accomplished by any one of a number of different means. It may be emphasised experimentally by ensuring that e.g. a touched object is touched successively from a number of different angles, or that a pursuing object takes more than one route to its quarry on different occasions.

¹³ I borrow this useful term from Crist (1999).

¹⁴ Though these distinctions need not, of course, be sharp. There is a distinction between red and pink even if there is no fact of the matter as to how certain intermediate shades are to be classified.

REFERENCES

- Bíró, S., and Leslie, A.M. (2007), “Infants’ perception of goal-directed actions: development through cue-based bootstrapping”, *Developmental Science*, 10, pp.379-98.
- Crist, E., (1999), *Images of Animals: Anthropomorphism and Animal Mind* (Philadelphia: Temple University Press).
- Davidson, D. (1982) “Rational Animals”, *Dialectica* 36 (4), pp.317-27, reprinted in Davidson (2001).
- Davidson, D. (2001) *Subjective, Intersubjective, Objective* (Oxford: Oxford University Press).
- Gelman, R. (1990), “First principles organize attention to and learning about relevant data: number and the animate-inanimate distinction as examples”, *Cognitive Science*, 14, pp. 79-106.
- Gelman, R., Durgin, F. and Kaufman, L. (1995), “Distinguishing between animates and inanimates: not by motion alone”, in: D. Sperber, D. Premack & A.J. Premack (Eds.) *Causal Cognition: A Multidisciplinary Debate*, pp. 150-84 (Oxford: Oxford University Press).
- Gergely, G. and Csibra, G. (2003), “Teleological reasoning in infancy: the naïve theory of rational action”, *Trends in Cognitive Sciences*, 7, pp. 287-92.
- Gergely, G., Nádasdy, Z., Csibra, G. and Bíró, S., (1995), “Taking the intentional stance at 12 months of age”, *Cognition*, 56, pp. 165-93.

-
- Golinkoff, R.M. (1983), "The preverbal negotiation of failed messages: insights into the transition period", in R. M. Golinkoff (Ed.), *The transition from prelinguistic to linguistic communication*, pp. 57-78 (Hillsdale, NJ: Lawrence Erlbaum Associates).
- Johnson, S (2000) "The recognition of mentalistic agents in infants", *Trends in Cognitive Sciences*, 4, pp. 22-8.
- Leslie, A.M. (1982), "The perception of causality in infants", *Perception*, 11, pp. 173-86.
- Leslie, A.M. (1984), "Spatiotemporal continuity and the perception of causality in infants", *Perception*, 13, pp. 287-305.
- Leslie, A. M. (1994), "ToMM, ToBY and Agency: Core architecture and domain specificity", in: L.A. Hirschfeld & R. Gelman (Eds.) *Mapping the Mind*, pp. 119-48 (Cambridge: Cambridge University Press).
- Leslie, A.M. (1995), 'A Theory of Agency' in: D. Sperber, D. Premack & A.J. Premack (Eds.) pp. 121-41, *Causal Cognition: A Multidisciplinary Debate* (Oxford: Oxford University Press).
- Leslie, A.M. and Keeble, S (1987), "Do six month old infants perceive causality?" *Cognition*, 25, pp. 265-88.
- Massey, C., and Gelman, R. (1988), "Preschoolers' ability to decide whether pictured unfamiliar objects can move themselves", *Developmental Psychology*, 24, pp. 307-17.
- Poulin-Dubois, D. and Schultz, T. (1988), "The Development of the Understanding of Human Behavior: From Agency to Intentionality", in: J. W. Astington, P. L. Harris & D. R. Olson, (Eds.), *Developing Theories of Mind*, pp. 109-125 (Cambridge, UK: Cambridge University Press).
- Premack, D. (1990), "The infant's theory of self-propelled objects", *Cognition*, 36, pp. 1-16.
- Spelke, E. S., Phillips, A., and Woodward, A. L., (1995), "Infants' knowledge of object motion and human action" in: D. Sperber, D. Premack & A.J. Premack (Eds.), *Causal Cognition: A Multidisciplinary Debate* pp. 44-78(Oxford: Oxford University Press).
- Spelke, E.S. and Van de Walle, G. "Perceiving and reasoning about objects: insights from infants", in: N. Eilan, R. McCarthy, & W. Brewer (Eds.), *Spatial representation*, pp. 132-61 (Oxford: Basil Blackwell).
- Woodward, A. (1998), "Infants selectively encode the goal object of an actor's reach", *Cognition*, 69, pp. 1-34.