Representation and Integration in Animal Minds

Peter Godfrey-Smith
Harvard University

A talk given at the "Comparative Psychology and Animal Minds" Workshop,
Harvard University, March 2011. Not published.

1. Frameworks

In attempts to understand the relations between human and animal minds, we often find ourselves making use of two models that derive from unpromising sources:

The belief-desire model, or folk psychology

I will sketch, but not argue for here, a hypothesis about its origins and structure. What philosophers think of as folk psychology has dual origins. One is a genuine "intuitive psychology." This is an evolved predictive tool seen also in some non-human animals and very young children. It is "peripheral" in what it recognizes and describes. Primarily, it recognizes *seeing* and *acting* (including trying) as activities of others. This intuitive psychology is a common element in how human and some non-human animals deal with each other.

The full folk psychology that we are familiar with in philosophical discussion also includes a model of psychological interactions within the periphery. This model is based perhaps on discourse, the exchange of sentences in public – roughly as Wilfrid Sellars

suggested many years ago. This part of folk psychology, unlike the first, might be expected to show some cultural variation.

The resulting framework might be fine for describing certain kinds of human thought, but it is problematic when applied more generally.

Associationism

Associationism had its origin in a "quasi-physics" of the mind, inspired especially by Newton. Locke developed an atomist view of the contents of the mind. Hume added to this a principle of "attraction" between these elements, akin to gravity. In the 19th century, J.S. Mill developed associationism on what he saw as chemical, rather than physical, lines. These early views have the aim of describing a set of interactions characteristic of mental objects, ways that a special set of objects act on each other in a mental arena.

Folk psychology and associationism were each shaped by the flow of data into more biologically reasonable versions. Associationism was transformed twice. Once was around the send of the 19th century. Rather than seeing the mind as an arena where ideas act on each other according to a quasi-physics, the brain was seen as a biological organ (Thorndike, Watson). This transformation included a shift to reinforcement learning, rather than classical conditioning, as the main mode of change. Reinforcement learning does not fit with the quasi-physics, but is plausible from the biological perspective. The mind (or just the brain) is seen as a biological pattern-recognition device that controls behavior. Associationism was transformed again about 40 years ago (as Celia Heyes described in her talk at this workshop), through the development of a less quasi-physical, more information-based, view of association itself; not the *proximity* of events but a *predictive* relationship between them is what matters.

With less of a clear lineage and without these two landmarks, folk psychology and its relatives have also evolved towards an information-processing approach that is less and less tied to a linguistic model of representation.

The two underlying models still affect discussions of human and animal thought, and often do so in their purer, less re-shaped forms. Associationism operates as a quasi-

physical causal story; mental dynamics are seen as due to association plus a set of motivational "drives." And folk psychology often appears in a form that has close links to the pattern seen in discourse.

This talk will emphasize cases that put pressure on familiar ways we have of drawing distinctions between human and non-human cases. The aim is not a squashing together project – a "they-can-do-all-we-can" view or a "we-are-simpler-than-we-thought" view. The overall picture I work within is one in which human language has "reprogrammed" our brains, in something like the way that people like Dennett, Spelke, and Carruthers argue. I will discuss that idea briefly towards the end, but much of the paper is about capacities of animals that show continuities with what are seen as important human capacities.

The first main section looks at some philosophical discussions of human/animal differences. This section is more argumentative. The second main section is more exploratory and more on the psychological side.

2. Self-control

In this section I will start by looking at contrasts made between humans and non-humans in some recent philosophical discussions. These discussions aim to admit some continuity between human and animal minds, but animals are also used as foils or contrasts to highlight some alleged features of humans that are of philosophical importance elsewhere. The discussions I will look at are due to John McDowell and Chris Korsgaard. Both philosophers discuss "autonomy" and a special kind of rationality as features of human action, which can be contrasted with the capacities of animals.

I will start with McDowell. In "Autonomy and its Burdens" (2009), McDowell wants to recognize both "continuity" and "discontinuity" between humans and animals (p. 193). The discontinuity involves a distinctively human capacity: acting on a reason *as the reason that it is*. This manifests rationality, and it "separates rational animals from other animals." This human capacity contrasts with acting on a reason in a weaker sense. When an antelope flees a lion, it acts on a reason, but does not respond to this reason – danger – "as the reason it is" (p. 192).

What is the difference between the two? McDowell appeals here to a metaphor of *distance*. In certain kinds of human action, we step back, we achieve some distance: "one must be able to step back, as it were, from the fact that a certain circumstance... inclines one in a certain direction" (192).

It is important to realize that this is a metaphor. It is linked to other spatial talk in philosophy, such as the idea of a "space of reasons." That second one I think that some people do not regard as metaphorical, but as literal description of something very abstract. Maybe, but the spatial talk used here, the "stepping back," is certainly metaphorical. It is imagined on the model of something a person might do with an external object like a painting. So it is natural to ask for something more than a metaphor here. McDowell does not say much more about what the psychology involved in action based on reason in this way is like, but he does say something about what makes it possible: "It seems plausible that this distanced orientation becomes a possibility with the acquisition of language" (p. 192). Without language, you cannot step back from inclinations, achieve distance from them. With language, you can. This makes humans rational in a richer sense.

My second example is Korsgaard:

> When I am aware, not just that I have a certain desire or fear, say, but that I am tempted to do something on the basis of that desire or fear, then it becomes open to me to step back from that connection and evaluate it: to ask whether my desire or fear provides me with a good reason to perform the action in question. And this enables me to take responsibility for what I do. This form of self-consciousness, I think, is what makes human beings rational and moral animals, and this is the one big difference that I have in mind. The other animals lead lives that are governed, I believe, by their instincts, desires, emotions, and attachments. Because we have the capacity to evaluate the influence of our instincts, desires, emotions and attachments on our actions, we are not completely governed by them. We have the capacity to be governed instead by normative standards and values, by a conception of what we ought to do. ("Facing the Animal you see in the Mirror," 2009)

As Korsgaard makes clear in *Self-Constitution* (2009, Chapter 6) "instinct" in her sense does not preclude a role for learning. Through learning, an animal can pick up on the practical significance of properties such as shape and color. This is a kind of intelligence.

But there is something extra humans can do, and this is again described using a metaphor of distance:

> Self-consciousness opens up a space between the incentive and the response, a space of what I call reflective distance. It is within the space of reflective distance that the question whether our incentives give us reasons arises. (*SC* Ch. 6)

In humans "instincts no longer determine how we respond to... incentives, what we do in the face of them. They propose responses, but we may or may not act in the way they propose." (*SC* Ch 6).

What is the role of these human/animal contrasts in Korsgaard and McDowell? McDowell says that he is not looking for too sharp a divide between the human and non-human cases. There are "intelligible precursors [of rationality] in ways in which perceived environmental circumstances engage with merely animal motivational tendencies." So precursors are admitted. And I do think that the description of human capacities given by McDowell and Korsgaard is picking up on something real. Most humans can, some of the time, reflect using conscious language-based thought on things they might do and why they might do them. They can consider the value of general principles as well as particular actions. Probably no other animal can do this. Let us assume that. And in making distinctions in an area like this, it is certainly possible to distinguish between (i) the fullest human capacity, and (ii) everything else. There can be good philosophical reasons for doing that. It then matters, however, what you say about what distinguishes the fullest-human cases from the rest. It is tempting to point to a sort of fundamental causal rearrangement when making such a distinction. Animals are "governed" by instinct and inclination, whereas humans can step back and establish a "space" between incentive and response. Some cases on the other side might be well contrasted with the human case in this way, but not all. The main example I will discuss here is self-control.

Animals are generally bad at self-control. With some animals the issue does not arise at all, but there are cases where it appears that a contrast can be drawn between a behavior that would exhibit self-control and one that would not, and the animal will choose the one that does not. We might see glimmers of self-control with well-trained

domestic animals, but it is hard to tell in informal settings what is going on. There is now experimental work on the topic though.

Here is one kind of self-control or "executive control" task, which has been done with both humans and other animals. You present the subject with two trays of some desirable food, such as candies, one tray with more than the other. The subject can choose a tray by reaching for it or pointing. But he gets the one he does not reach for. So if there is more food on tray 1 than tray 2, in order to get the larger amount he must reach for or indicate the smaller amount; he must take tray 2 to get what is on tray 1.

The first studies were done with chimps, by Sarah Boysen and her co-workers.[1] The results were enormously suggestive. The chimps Boysen worked with could not learn to indicate the worse option to get the better one, despite hundreds of trials, and despite evident frustration on the part of the chimps. But these particular chimps had learned to use arabic numerals to refer to quantities in other experiments. When the actual candies were replaced by numerals, and the chimp had to choose the smaller numeral to get the candies corresponding to the larger one, she could do it. The result suggested something about the usefulness of arbitrary symbols as tools for representation, making it possible to overcome simple compulsions and drives.

Later work revisited the initial result that chimps were unable to choose the smaller to get the larger quantity, without the aid of symbols. This is a hard task for non-humans to learn. But some can do it; they can learn in reasonable time to overcome the tendency to indicate the larger prize. Orangutans succeeded, some individual chimps did, and sea lions did best of all.[2]

The orangutan result has been questioned because the orangutans did not show an initial preference for the larger quantity, so they did not have to overcome this response.

---

[1] See Boysen and Bernston, " Responses to quantity: perceptual versus cognitive mechanisms in chimpanzees (*Pan troglodytes*)," 1995. Boysen, S.T., G.G. Berntson, M.B. Hannan and J.T. Cacioppo (1996) "Quantity-Based Interference and Symbolic Representations in Chimpanzees (*Pan troglodytes*)," *Journal of Experimental Psychology: Animal Behavior Processes*, 22: 76–86. Boysen, S.T., K.L. Mukobi and G.G. Berntson (1999) "Overcoming Response Bias Using Symbolic Representations of Number by Chimpanzees (*Pan troglodytes*)," *Animal Behavior and Learning*, 27: 229–235.

[2] Gentry and Roeder (2006),"Self-control: why should sea lions, *Zalophus californianus*, perform better than primates?"

In another self-control task orangutans did better than all other apes, however, tested on exactly the same task. This task involves reaching for something in a situation where the most direct reach merely knocks the food away, and an indirect reach is needed to get it. Orangutans here did better than 3-year old humans, and about as well as 4-5 year old humans (Vlamings, Hare, and Call 2010). So there does seem to be something in the idea that orangutans are good at self-control. Putting sea lions and orangutans together, we have here a glimpse of a pattern that goes against the influential idea that complex social life makes for more sophisticated intelligence. Orangutans are less social than chimps and other great apes. Sea lions do not have a complex social life though they do have harems. It is suggested that animals less caught up in hurried contests for food, as chimps are, can do better because they can – to use the spatial metaphor – learn to step back a little from the immediate attractiveness of the food before them.

The "social intelligence hypothesis" has a good deal of evidence for it. But against that background, this is an interesting pattern in a different direction. Those of us who were a little unenthusiastic about the rise of the social intelligence hypothesis because we value solitary reflection can take heart.

These results are important because in a case like this the animal is definitely after food, something it does not need to learn to want. A sea lion most definitely wants lots of fish. But the way it can get lots of fish is not by taking them, but instead by suppressing that fish-directed impulse. If you want to use a spatial metaphor, the animal here "opens up a space between the incentive and the response." This spatial metaphor, as noted earlier, seems most applicable when there is linguistic representation or something like it. We humans can consider, deliberately, a sentence, by saying it to oneself. I don't think the sea lions are doing that. But they are doing something that involves a partial overcoming of instinct and inclination.

Returning to the contrasts between humans and animals drawn by McDowell and Korsgaard: human capacities for choice are undoubtedly distinctive in philosophically important ways. Language probably has a lot to do with this. But in discussions in which freedom and rationality are what the philosophers have their eye on, the human/animal contrast often takes a particular form. The contrast between the full-human case and others is made by attributing a kind of passivity in the face of instinct to everything on

the non-human side. If a subtle enough concept of activity or rationality is employed, the distinction can be made that way. But then the cases on the non-human side will include self-control, and the ability to suppress initial responses in the service of better handling of a situation and its possibilities. The cases on the other side will include a lot that we might have thought of as the *antithesis* of a passive or instinctive or impulsive response.

I will mention one other case. In the experiments discussed so far, a non-human is able to achieve some distance from a motivational state. One view of what is important here is an ability to think about one's own inner states. If that is what is important, there are results that bear more directly on it, not with respect to motivational states but with respect to *memory* – a belief-like state. Robert Hampton tested rhesus monkeys on the following task.[3] They first saw an image and, after a delay long enough to make the task difficult, had to re-identify the image from a range of possibilities. Then a choice was introduced before the memory test. Before the testing phase, the monkey could "opt out" of the test, guaranteeing it a low-quality reward, or opt to take the test, in which case it got a large reward if its answer was right and no reward if it was wrong. This choice was offered in 2/3 of cases. In the other 1/3, the animal was forced to take the memory test. The idea was that if the animal does better on the tests it chooses to take than on the ones it is forced to take, this indicates it has some ability to assess its own state – it can work out whether has remembered the image well enough to make the test a good bet. Hampton found that the monkeys did indeed do better on chosen than forced tests. Pigeons, in contrast, did not show this ability. If the spatial metaphor has not run out of steam here, then this is another case of a non-human establishing some "distance" from some of its own psychological states.

3. Integration

As Peter Carruthers outlined in his talk, it has been suggested that human cognition makes use of two kinds of processing. "System 1" is evolutionarily old, parallel, unconscious, and does not conform well to normative theories of rationality. "System 2"

---

[3] Hampton, "Rhesus monkeys know when they remember," *PNAS* 2001.

is newer, conscious, and more rational in its operation. System 1 has often been seen as associationist in nature, though Carruthers does not see it that way. A functional rationale for this division is sketched in this quote by Evans and Frankish.

> The advantage of the dual-process system is that conscious reflective thought provides the flexibility and foresight that the tacit system cannot, by its very nature, deliver. Most of the time our decision making is automatic and habitual, but it does not have to be that way … consciousness gives us the possibility to deal with novelty and anticipate the future. (Evans and Frankish, *In Two Minds*, 2009)

Early versions of this view posited or seemed to posit two sets of machinery in humans, passing information back and forth. Animals don't have the second system and are stuck with the less rational, perhaps associationist, system 1.

The idea of two machines in humans, passing information back and forth, was perhaps always intended as a simple first sketch. More recent work, like that of Carruthers, is trying for a better picture of the relation between two sets of capacities here.

I think there might be something important in the dual-process view, but existing discussions may be too influenced by bifurcations coming from the old choices between causal models in this area, of the kind outlined in the first section of this talk. There are ways in which animals can achieve some system-2-*like* capacities. This might lead us to reject the dual-process model altogether, as a too-sharp partition of a more complicated mixture of capacities, or it might leave room for a modified dual-process view.

I will approach this from a particular direction, through questions about the *unification* of cognitive processes. Discussions of unity have a philosophical history, featuring Kant's response to the fragmented picture given by Hume. I will treat the issue in a functional way. There is a general "design choice" relevant to the evolution of behavior controllers, a choice between having integrated control as opposed to several separate causal or information-processing streams. This is related to the distinction between serial and parallel processing. This issue is often now discussed in terms of "modularity" hypotheses. Here I will look at a simpler, more experimentally accessible

aspect of disunity. This is *lateralization*, the specialization of halves of the brain, in vertebrates, for different tasks.

There are simple and interesting results here. A variety of species appear to be more reactive to predators seen in the left, rather than right, side of their visual field.[4] Given the crossover of neural pathways, this means that the right hand side of the brain is more involved in scanning for predators. A preference in several kinds of fish for an individual to position itself so that the image of a conspecific is on its left side has also been reported. Even tadpoles have been shown to prefer to position themselves so that the image of a conspecific is on their left side. On the other hand, rightward biases for prey catching and foraging are seen when the prey or food has to be discriminated from similar targets, so this gives a special role to the left side of the brain. Rather than lateralization being a quirk of the linguistically affected human brain, revealed mostly in unusual circumstances such as split-brain cases, other animals have invested greatly in lateralization, and there is less traffic across the two halves.

Why? This sort of thing seems to have clear disadvantages. It seems to leave the animal vulnerable to attack on one side, or less able to find prey on one side. Two related and compatible hypotheses, both quite intuitive, have been offered.

1. Different tasks involve different styles of processing, and the brain has specialized each hemisphere for one kind of task.
2. Lateralization is good when you are multi-tasking, doing two things at once.

(See Vallortigara and Rogers for both ideas).

In a test of the latter possibility, Lesley Rogers and others were able to raise chicks with reduced lateralization of their brains – not with surgery, but by altering the light conditions in their development. They were then tested on various tasks, including a combined task: both finding food amid pebbles that look similar, and responding to the image of possible predator flying overhead. Less-lateralized chicks did very badly at the combined task. This was not, the authors say, because of general deficits.

So a lack of integration has some advantages. But so does integration, in other situations. An obvious thing to think about here is the problems of dealing with novelty.

---

[4] See Vallortigara et al. 1999 for a review.

The dual-process advocates make much of this when discussing the role of system 2, as seen in the quote from Evans and Frankish I gave earlier. Animals have ways of achieving some degree of integration despite an architecture that makes it less straightforward. Staying with the case of birds, most bird eyes are on the sides of their head, with a small area of overlap of visual field, and with the two retinas feeding information to different parts of the brain. Birds do not have a large bridge across the two halves of their upper brain like our *corpus callosum*. There seems normally to be very imperfect integration of information across the two eyes. In a pigeon experiment, birds were taught a task using one eye only, and then tested on the task with the other eye. There was very little transfer.[5] There was even poor transfer between the monocular and binocular regions of *one* eye.

In the light of this, here is a simple finding by Marion Dawkins.[6] She studied head movement in chickens. She found that that hens tend to approach a novel object in a distinctive way, not applied with familiar objects. They approach a novel object moving the head around so that all parts of both eyes are exposed to it, switching between front-on and side-on viewing and moving their head from side to side. In a bird of this kind, the only way for the whole brain to get the information is to feed it through both eyes, and all the parts of both eyes. It seems that the weaving gaze of a bird is designed to slosh the incoming information around, to make sure it is "broadcast" to all parts of the system. This is a bit of adaptive integration, directed at handling novelty, achieved through behavior.

This relates also to phenomena of *attention*, the directing of extra resources to one aspect of a present scene or situation. This I do not know so much about. But some degree or analogue of attention is seen in many animals. In the human case, attention has a fairly clear cognitive role in dealing with novelty and enabling deliberate action (Dehaene and Naccache 2001).

I will finish by moving from these discussing first hints of cognitive integration to what might be its culmination, linguistically based thought.

---

[5] Ortega et al, 'Limits of intraocular and interocular transfer in pigeons'.
[6] Dawkins, 'What are birds looking at? Head movements and eye use in chickens'.

There has been a surge of recent work on the cognitive role of language. Here I will put ideas together from many discussions – Dennett, Carruthers, Spelke, and Kritika Yegnashankaran's dissertation, written in this department – and discuss two apparent cognitive roles of language that are relevant here.

1. Language is a medium that can be used to combine information of disparate kinds. In Liz Spelke's version of the view: language is a medium for bridging self-contained "core knowledge" systems that operate usually independently and in parallel. Language "allows representations to be combined across any cognitive domains that we can represent and to be used for any tasks that we can understand and undertake."[7]

2. Language is a medium that lends itself to the "broadcasting" of information to many parts of a cognitive system. This is a psychological feature that derives from the primary social, interpersonal uses of language. Language is both an input and an output system. We both make sounds and interpret those of others. This combination lends itself to *intra*-personal use of a special kind as well.

This is discussed in detail by Carruthers. Credit might be given to Vygotsky for a first sketch. Stripping away some details of Carruthers' view, the idea is that the evolution of language already requires some way of bringing the outputs of parallel systems together so that all could receive expression in speech. Once this is happening, this provides the basis for a feeding of sentences, even unspoken, back through the *input* end of the system, in such a way that a "heard" version of the sentence is produced in auditory imagination. The sentence can then be used by many parts of the system in further processing. So by means of inner speech, information can be combined, organized, and psychologically broadcast.[8]

---

[7] "Natural languages provide humans with a unique system for combining flexibly the representations they share with other animals." (2003, p. 291)

[8] "By virtue of being “heard,” then, the sentence would also be taken as *input* to the conceptual modules which are down-stream of the comprehension sub-system of the language faculty, receiving the latter’s output. So the cycle goes: thoughts generated by central modules are used to frame a natural language representation, which is used to generate a sentence in auditory imagination, which is then taken as input by the central modules once again." (2002)

These two features of language have analogues in the non-human case, and the distinctive contribution made by language can illustrated by looking at the analogues.

Seyfarth and Cheney, in *Baboon Metaphysics*, argue for a "language of thought" in baboons. The argument is that even though baboons' vocal *production* capacities are very limited, their *comprehension* capacities show they have a powerful internal system of representation. Though each baboon vocalization does not have significant syntax or combinatorial power, a series of vocalizations in rapid succession made by several unseen animals can carry a very complicated message. The series can carry information about changes in dominance rank, about the mending of a dispute or its exacerbation. There is a lot of information present because although baboons can make only a few calls, which have no significant syntax, they can recognize who made a given call. A series of calls by specific individuals can then have a kind of implicit syntax – that is, understanding what it all means requires skills comparable to those that would be required to understand a single syntactically complex call. Recording experiments show that baboons can respond to this implicit syntax. For example: a threatening call by a subordinate followed by a submissive call from a dominant is processed as a surprising sequence, in a way reflected in behavior. The converse is not. Cheney and Seyfarth say that this shows that baboons have a "language of thought," because their processing of vocalizations shows an ability to respond appropriately to complex combinations, including novel ones.

Liz Camp has looked closely at this argument. She thinks it does not support a *language* of baboon thought, though it might support the presence of some other representational system. What is the difference? Camp emphasizes the expressive power of language. Any powerful representational system includes a "combinatorial principle" and a "referential relation" – something that assigns meanings to simple signs.

> In language, this [combinatorial] principle is highly abstract, and has a very thin semantic significance. Predication, for instance, signifies instantiation or property-possession, so that the significance of combining "Socrates" with "is wise" is that the referent of the former instantiates the property expressed by the latter; and this relation is sufficiently abstract and general that it can relate nearly any property and object. Further, in language the referential relation mapping basic expressions to objects and properties in the world is

> conventional or causal. Taken together, both the referential relation and the combinatorial principle are abstract enough that they don't impose substantive in-principle limitations on what can be assigned as referents to those basic expressions. (Camp, "A Language of Baboon Thought?")

Language is an exceptionally flexible medium for representation, because of these features of its format. There is no reason to think the baboons have something like this. You can call what they have a "language" if you like, but it might have more in common with a map-like form of representation.

We can add to Camp's points with the aid of Carruthers' view. This inner system is apparently not linked to a capacity for internal "broadcast." Perhaps it is, in some hidden way, but not in the way that language works, which is by auditory imagination of complex signs.[9] If the impoverished nature of the "production" side of baboon vocalization applies also to anything done with those signs internally, then there is not yet any reason to think baboons can engage in the free production of complex inner signs, even though they might know what to do with them if they had them. All we have presently is an argument for believing in is a capacity for some kind of combinatorial or system of representation that acts in *response* to complex input. This does not necessarily bring with it the capacity to create *new* representations of the relevant kind. This capacity seems likely to be of special importance in the kind of decision making that people like McDowell and Korsgaard are trying to understand, discussed in the first main section.

So this last point connects back to those in the second main section. There I argued that the categories we use in making human/non-human distinctions lead, in effect, to an underestimation of the resources of animals. Here I am arguing that they may also lead to an overestimation. The primary aim of Cheney and Seyfarth's argument here was to oppose an associationist view of the baboons. Their argument works, I think, with respect to that goal. They are right that associationist explanations – of at least the forms they have in mind – seem not to apply here. That leads Cheney and Seyfarth to a ready

---

[9] Broadcasting of this kind is a bit like what Marian Dawkins's chicks are doing: deliberate sending of information to the disparate parts of the system, overcoming cognitive separation through action of the whole organism.

acceptance of a "language of thought" view that would bridge the human and non-human case. But languages of the kind seen in humans may still add something substantial to the minds of those that use them.
