Commentary on three papers about animal cognition

Jonathan Bennett

From *The Behavioral and Brain Sciences* 1 (1978), pp. 556–560.

1. Some remarks about concepts

Comments on E. Sue Savage-Rumbaugh, Duane M. Rumbaugh, and Sally Boysen, 'Linguistically mediated tool use and exchange by chimpanzees (Pan troglodytes)', *Behavioral and Brain Sciences* 1 (1978), pp. 539–554.

Savage-Rumbaugh et al. show that chimpanzees can interact, in intentional ways, using symbols. It is good to get away from interactions between chimpanzees and humans The mere fact that one participant is human generates a bottomless reservoir of possible contributions to the chimpanzee's performance; these are all filtered out when the chimpanzees are induced to interact with one another.

It's also valuable that the chimpanzees faced problems which, though they were set up by humans, owed none of their essential features to that fact. All the problems had the form 'How can it be opened?' rather than 'What does E want me to do?' In moving from essentially contrived problems to possibly natural ones, SR&B have escaped from speculations about the chimpanzee's beliefs about the trainer's desires (see my comments below on P&W); and that filters out a further possible source of perturbation.

SR&B are to be congratulated on this elegant experiment. Unfortunately, their interpretation of it is flawed by their uncritical use of certain concepts. Their work breaks new ground, they say, because it 'has demonstrated that two chimpanzees have been able to comprehend the symbolic and communicative function of the symbols they use'. But if they have a definite sense for the word 'communicative', it is one which makes that claim obviously false; and I cannot find any one sense that they are giving to the term 'symbolic'.

SR&B use 'communication' to mean something distinct from—perhaps narrower than?—problem-solving, and this generates for them a criticism of Premack: 'Does Sarah give evidence of comprehending that she is communicating with her teachers, or that they are communicating with her, as opposed to simply solving a set of problems?' One would like to know what distinction SR&B have in mind here. This matters, not just for their importance-claim, but also for the following reason.

The performances of SR&B's subjects are analogous to injunctions—requests. commands, pleas, etc.—which aim to elicit behavior from the other party. Contrast injunctions with statements, which aim to produce belief (or awareness, or realization, or knowledge) in the other party. Now, what is wrong with saying that an animal which purposively does anything is thereby requesting or commanding the universe to produce the desired result? If there were nothing wrong

with that, then all purposive behaviour would be the uttering of injunctions, and all practical problem-solving would be 'communication'. There does exist a stop-gap answer to this, namely that we always restrict 'communicate' to cases where both parties to the transaction are sentient. That is surely right. But why do we want this restriction? If it is not a mere aesthetic preference, a superficial linguistic nuance, then there must be some underlying rationale for it, some further strand in the concept of communication which will lie idle unless the communicators are both sentient. What can this further strand be? One possible answer is this: 'Unless both parties to the transaction are sentient, the communicator can't produce a belief in the other; and the latter is involved in all communication.' That is not the only nonarbitrary way of stopping the notion of communication from flattening out so that it covers all purposive behaviour; but it is one plausible way, and I think it is the one SR&B would choose. They repeatedly appear to assume that communication essentially involves transfer of information, this being understood as production of belief or awareness or the like. This implies that the basic purpose of an injunction such as 'Hand me the stick' is—in SR&B's phrase—'the transfer of information regarding the necessary tool': that is, the command is really the statement 'I want you to hand me the stick'.

I like that account of injunction, and the account of communication in general which goes with it. I think that in most human communication the speaker does intend to produce in the hearer a belief-change which may, but need not, be intended to have some specific further behavioural upshot. That's over-simplified (see Bennett, 1976, §41), but it will do for now. Simplified as it is, it carries an impressive conceptual load: X *intends* to make Y *believe* that X *wants* such-and-such. And on any viable account of intention, that involves: X *believes* that X can make Y *believe* that X *wants* such-and-such. SR&B write as though they were content with this. They find it 'difficult to understand how [Premack's] Sarah could come to realize that the plastic chips could be used to communicate desires', and they apparently think that *their* chimpanzees did come to 'realize' such things. On the most modest construal of this that I can devise, it involves a belief about a belief about a desire. See also their pregnant remark: 'The initiator expected the recipient to understand why he was gesturing.'

Although this is the concept of communication which SR&B manifestly employ, they couldn't be content. with it if they saw what it involved. For they would see that their experimental work goes no way towards showing that chimpanzees can 'communicate'—let alone believe or comprehend that they are 'communicating'—in this strong. sense.

It's puzzling that one should have to say these things. SR&B themselves write 'The question of whether signing chimpanzees comprehend the nature, function, and symbolic power of the symbols they use becomes a question of awareness and intentionality'; and evidently they see this as a problem area which is 'now being reopened'. I can't understand how, if they realize *that*, they can have permitted themselves to use the concepts of awareness and intention in such an innocently uncritical manner, both in evaluating their own work and in criticizing that of others.

Now let us turn to 'symbolic'. SR&B do not offer to explain this, and it isn't self-explanatory. Sometimes they apparently use it simply to mean 'non-iconic': but then what does 'iconic' mean? I would explain it in terms of *natural* associations between features, in contrast to those that are *institutional*, *artificial*. *stipulated*, and so forth; but then one could hardly credit chimpanzees with making a distinction between icon and symbol—with 'understanding that they could make symbolic requests to one another' or

'comprehending the symbolic nature of the materials they are using'—at least not on any evidence presented in this paper. It looks as though SR&B would tie 'icon' to the notion of something that resembles what it stands for, and that generates a feeble sense of 'symbol' in which the chimpanzee probably can distinguish icons from symbols, that is, similar pairs of items from dissimilar pairs.

But neither of those senses of 'symbol' makes any sense of SR&B's crucial use of the term in their importance-claim: their chimpanzees. they say, comprehend the symbolic function of the symbols they are using, and elsewhere items are said to have symbolic *power*. These phrases are not explained at all, and couldn't be explained by reading 'symbolic' as 'non-iconic' on any sane account of what 'iconic' means.

An item's symbolic function (or power) might be its use (or capacity to be used) to symbolize or stand for something else; but SR&B surely do not think that their results show their chimpanzees to have a grasp of the relation of *standing-for*. The chimpanzees do, it is true, know that certain symbols are *associated with* other things, but other chimpanzees have been shown to know that much. They know further that these associations can be exploited for practical purposes, but there's nothing new about that either.

The best guess I can make is that in the phrases 'symbolic power' and 'symbolic function' the word 'symbolic' is being given the same meaning as 'communicative'. SR&B do sometimes write as though they took those two terms to be equivalent. Thus they write: 'It is impossible to tell whether the chimpanzee is simply imitating or echoing. . . the action or object, or whether the animal is indeed attempting to relay a symbolic message.' Here something which is both symbolic and communicative is contrasted with something which is neither, and there is no attempt to sort out the ingredients of the mix; so the suggestion is that '(non)symbolic' and '(non)communicative' must stand or fall together. The same thing is firmly implied when SR&B allude to 'the symbolic and communicative function'—not *functions*—'of the symbols they use'.

Of course I do not accuse SR&B of not knowing that the terms 'symbolic' and 'communicative', far from being equivalent, are perfectly independent of one another. I'm sure that they do know. But I think that they have slipped into a way of writing and thinking which belies this knowledge. If I am right, then what seems to be a double importance-claim is really the single claim that the chimpanzees knew that their behaviour was communicative. If I am wrong, and SR&B's 'symbolic function' and kindred phrases mean something other than 'communicative function', I am at a loss to know what that meaning can be.

2. Beliefs about beliefs

Comments on David Premack and Guy Woodruff, 'Does the chimpanzee have a theory of mind?', *Behavioral and Brain Sciences* 1, (1978) pp. 515-526.

If one wants to apply to chimpanzees a concept of communication as rich as SR&B's, one must first ask what evidence there is that chimpanzees can have beliefs about beliefs about *anything* (let alone beliefs about beliefs about desires). It is good news that P&W are investigating the question in a properly isolated way, getting belief-about-belief out of the undergrowth and into the open where it can be looked at squarely. The work—achieved and projected—is so interesting and potentially important that I would do anything to be allowed to help it along. I here offer what I can: three criticisms, a comment, and a suggestion.

First, a small point which illustrates a large one. The 'empathy' interpretation of the first experiment is said to 'assume that the animal imputes purpose to the human actor, [but] does not grant the animal any inferences about [the human's] knowledge'. That is wrong, I submit. Sarah's behaviour isn't relevant to what she thinks the human *wants* or *intends* unless she thinks that he *thinks* (perceives, cognizes) his situation to be thus and thus.

The large point is that when one is attributing mental states on the evidence of behaviour, cognition and motivation must go hand in hand. An animal's actions show what it thinks only on certain assumptions about what it wants, and they show what it wants only on assumptions about what it thinks. This goes for our attributions to the chimpanzee, and for hers to the human: if she is perfectly agnostic about the human's cognition, how could she have any thought about how his purposes would lead him to behave, or about what purposes his behaviour would manifest? If she can have neither, then she can have no thought about his purposes.

So I dissent from P&W's conjecture that 'inferences about motivation precede those about knowledge'. If the inferences are to be based upon behaviour, it is *impossible* for there to be this separation (see Bennett, 1976, § 15). This is worth stressing, if it is true, as a warning against devising experiments where one of the two factors is attended to while the other hovers in the background, unrecognized and thus uncontrolled.

Secondly, some of the experiments are, as P&W recognize, disturbed by the question, What is Sarah up to? Presumably she *wants* to succeed in her assigned task, but that is an empty conjecture unless we know what she thinks her assigned task is. Although P&W are properly cautious about this, the situation may be worse than they recognize. If all goes according to plan, Sarah thinks, in the first experiments, that she is to *predict* what the human actor will do; in the next lot she thinks she is to *express what she prefers*; and in the 'embedded videotape' experiments she thinks that *she* is to predict and that *he* is to express what he prefers. I'm not optimistic.

Might it not be better to escape from experiments whose interpretation depends on what the animal thinks its assigned task to be? My objection is not that that concerns what she believes the trainer wants, and thus concerns her psychological theory, for we can surely construe 'what she thinks the assigned task is' so that it doesn't automatically credit her with any psychological theory. The objection is just that this feature introduces a not easily remedied uncertainty about how the results—whatever they are—should be interpreted.

Thirdly, although P&W's current experiments on the attitudes of chimpanzees to liars and fools have the merit of not involving the concept of an assigned task, they do strike me—as clearly they strike P&W also—as being highly tenuous for other reasons. I hope that more work will be done on the rock-bottom matter of belief about belief before much more effort is expended on these more complex and recherché matters—in which I include the distinction between knowing and guessing. I don't deny that these matters are amenable to experimental investigation, but I suggest that they are better left aside until the groundwork has been more completely done.

Fourthly, as P&W imply in their closing remark, the price of avoiding mentalistic concepts is an increase in the complexity of the conceptual structures needed to handle the materials, and that is what justifies using mentalistic concepts. Now, suppose we know that in the first experiments Sarah was predicting how the human actor would behave, and let us ask: Why shouldn't she have reached

this prediction by purely behavioural computations, without going through intermediate stages concerning his states of mind? P&W answer this, it seems, by saying that if her route to the prediction wasn't mentalistic, it must be describable in 'associationist' terms, and they object to this for reasons that I am not sure I understand. Perhaps what they are saying comes down to what I contend to be the best answer (if it is true), namely: For Sarah to get from data to prediction by a route which doesn't attribute beliefs to the human actor, she would need an extremely complex inference, whereas she could get there without undue complexity if along the way she had hypotheses about the human's mental states. That claim about relative complexity is needed to justify crediting Sarah with beliefs about beliefs and not mere beliefs about behaviour.

Finally, whether the relative-complexity claim can truthfully be made in connection with those first experiments of P&W's is not clear to me, because I don't know enough about what Sarah's premises were. Exact information about an animal's data would be easier to get if the experiment had to do not with 'Does she think that he knows that heaters have to be plugged in?' but rather with 'Does she think that he knows that this heater is not plugged in?' That is, there would be better grounds for a belief-about-belief interpretation if the focus were on her beliefs about what the human believes about this particular situation. For then one could vary the evidence she had about his knowledge of individual features of the situation—features which were relevant to whether or how the problem could be solved. Let her see that the key is turned while he is not looking; or that the box is sneaked away after being hidden from him by a screen; and so on. This could provide strong evidence for beliefs-about-beliefs, by showing that the chimpanzee must be making inferences which are dauntingly complex when stated in purely behavioural terms but relatively simple when stated in terms of mentalistic concepts.

How is she to manifest these predictive beliefs about the other party's behaviour? I suggest the following. Let A be the agent about whose beliefs we hope the subject chimpanzee will form beliefs. Give the chimpanzee abundant evidence of A's sharing her value-system and being prepared to cooperate with her towards common ends. This is to justify us in assuming that if she has any psychological theory about A, its motivational part will credit A with roughly the motivations that she herself has. Then construct coordination-problem situations, where the chimpanzee can see that *what it is prudent for her to do* depends upon *what A is in fact going to do* (see Lewis, 1969, ch. 1). Her predictions about A will then be manifested in behaviour which can be much more confidently interpreted than could any amount of photograph-selection—namely, in terms of her pursuit of her own natural down-to-earth goals. I don't mean that interpretative problems may not still arise, but merely that they aren't likely to be problems about what her motivations are, what she is 'up to'.

I cannot put experimental flesh on these bones: no doubt Herculean labours are needed to turn such armchair proposals into a practicable programme. Still, abstract proposals can have value, and I offer mine, hopefully, as pointing to a possible kind of experiment which **(1)** keeps both belief about motivation and belief about cognition clearly in view; (2) avoids the notion of 'the assigned task'; **(3)** can be aimed at the most elemental kinds of belief about belief, leaving the complex ones until later; and **(4)** makes it possible to know fairly exactly what the animal's data are, and what her conclusion is, thus facilitating the comparison (in respect of complexity) between the mentalistic and the non-mentalistic routes from the one to the other.

3. The need for criteria

Comments on paper by Donald R. Griffin. 'Prospects for a cognitive ethology', , *Behavioral and Brain Sciences* 1 (1978), pp. 527–538.

I share G's interest in developing a richly mentalistic psychology and ethology, but his advocacy, persuasive as it is, would be stronger still if he were more cautious in his deployment of mentalistic concepts. That would involve two things.

(1) More clarity and explicitness are needed regarding the criteria for applying mentalistic concepts. G says a good deal about this in particular cases, for example, in his good remarks about the possibility that honey-bees might learn to report on fiberglass; but his campaign needs a general statement as to what sorts of behaviour support which mentalistic attributions. It would be controversial, and would need defence, but the mere statement of it would give us a better idea of what G wants us to accept.

That statement of criteria should govern the discussion. One feels the need of it in, for instance, G's remarks about chimpanzee-communication experiments. He seems to assume that some of those studies provide evidence of 'communication of intentions' in some significant sense, unless all of them are vitiated by 'Clever Hans' errors. But there are hosts of ways in which those experiments might, without being vitiated by cueing, have to be interpreted as something other than communication of intentions.

The desired statement of behavioural criteria would direct more attention to the internal geography of our system of mentalistic concepts. G tends to cut corners, as in his handling of the case of the two planes, one guided by a kamikaze pilot and the other by a heat-seeking device, where 'communication. . . might well convince us that a real, live, conscious pilot was flying one machine'. So indeed it might, but how and why? We are to be impressed by the pilot's responding appropriately to our messages; and presumably the 'simple dialogue' mustn't be so stereotyped as to revive the suspicion that we are talking to a tape—for example it mustn't consist just in commands from us and 'Yessirs' from the pilot. So the dialogue is to provide evidence that we should get appropriate replies to most things we might say; and the only thing capable of such a feat is a human being—a mind-endowed, conscious, aware person. Now, I don't dispute that a simple dialogue could convince me that the plane was piloted by a conscious being. But two corners have been cut.

Firstly, the story is not essentially one about communication. It's true that we are in doubt about the pilot until we discover that he can communicate with us; but that is a large, cloudy truth which contains within it the leaner and more precise truth that we are in doubt about the pilot until we discover that he can respond appropriately to a large range of input. The input is provided by us, and could in that sense be called 'signalling', and thus 'communication'; but that is a mere accident of the example, and not part of what makes the dialogue evidence as to the pilot's status.

Secondly, G's route from 'The dialogue is succeeding' to 'There is a conscious being at the other end' is invalid, for a reason which has nothing to do with whether 'communication' is involved. The dialogue is evidence that the plane's guide would respond appropriately to a vast number of distinct signals, but that is not *in itself* evidence of consciousness. It points to consciousness only with the help of the fact that in our region of space-time the only things which have such a large store of responses are the higher animals, and they are all conscious. What qualifies them as conscious, however, is not the size and complexity of their stores so much as their flexibility, adaptability, educability. It is true that in the animal kingdom complexity and flexibility

tend to go together, with humans having a uniquely high degree of each; and this is no mere coincidence. Still, they are distinct properties of an organism, and they are not equally relevant to consciousness or mentality. If there were a tremendously complex hard-wired device which could handle the pilot's end of the 'simple dialogue' but which was perfectly ineducable, it would pass G's dialogue test yet wouldn't be conscious (see Haugeland 1978). I'm not saying that G's test isn't good enough ('Maybe it's not a man answering us, but rather a technologically innovative device. . . '). I'm saying that he doesn't make sufficiently clear why it is good enough.

(2) When the criteria are actually formulated, they should not be too generous, i.e. they should not make it too easy to establish that a given animal is conscious. This is not because consciousness should not be spread too widely: I hold no brief for the view that we must keep the other animals at a respectful distance from ourselves. Rather, the reason is that the more lax the criteria for something's being X are, the less content there is in the statement 'This animal is X', and so the less interesting such statements are. Sometimes G tends to push towards making such statements true at the price of making them uninteresting. For example. in one place he asks plaintively 'what else bees might be expected to do that would provide stronger evidence of intention to communicate, given the circumstances under which their behavior has been studied so far'. This mislocates the onus of proof. Until we have good evidence that honey-bees intend to communicate, we should not say or think they do; otherwise our uses of such expressions as 'intend to communicate' become empty and boring. **(**My own guess is that the evidence will never be forthcoming, and I wish G hadn't permitted himself the jibe 'Is it because bees are small?', ignoring a reason which surely deserves a respectful hearing—namely that the neural organization of bees may be too simple to permit that adaptability and flexibility which many of us regard as criterial for mentality. But that's by the way.**)**

G's remark about bees' intention to communicate, having served to illustrate my main point about severity of criteria, also illustrates my general theme. See what precedes it:

> Smith [says] that honey-bee dances communicate information 'about characteristics of the next flight the dancing communicator will make' rather than about the location of something desirable. But the distinction between predicting one's future behavior and expressing an intention is a rather subtle one that is certainly difficult to analyze in another species. It is therefore appropriate to ask what else bees might be expected to do that would provide stronger evidence of intention to communicate. . .

This contains two conflations. Firstly, it conflates **(i)** the difference between flight-prediction and report-on-food with **(ii)** the difference between flight prediction and report-on-flight·intention. It will be hard to get any purchase on **(i)** in the apian context, and G is right that **(ii)** is elusive in any context; but he writes as though they were the same distinction, when in fact they are as different as chalk from cheese. I can only suppose that this conflation results from a kind of conceptual hurry that appears to be present throughout the article.

Another sign of hurry occurs in the next transition in the quoted passage. There, G glides from 'expressing an intention' to 'intention to communicate', as though he didn't distinguish 'X **communicates that it intends** to fly' from X '**intends to communicate** that it will fly'. But these two are also quite different: the question of what is communicated, for example whether it is a message about intentions, is independent of whether the communication is intentional.

Just because I find G's campaign so sympathetic, and so many of his details interesting and persuasive, I would like to urge upon him the importance of circumspection—of a patient, careful, continuous attention to conceptual foundations.

References

J. Bennett *Linguistic Behaviour*. Cambridge Univ. Press. 1976.

D. K. Lewis, *Convention: a philosophical study*. Harvard University Press, 1969.

J. Haugeland 'The Nature and Plausibility of Cognitlvism', *Behavioral and Brain Sciences* 1 (1978), pp. 215–226.