Note: While in Australia the past four weeks, I talked with Andy Richards and John Pepper, read a chapter in Radner & Radner that John was reading and several articles in things I picked up at airports (especially Seyfarth & Cheney and the Time article on animal Intelligence), and ended up writing the following essays in my journal. They seemed pertinent to the seminar, so, even though they are really two essays, I combined them here. Please realize that this is a first draft with not much of a look at the literature, so it's crude.

[Varieties of animal signals and a theory of mind]

Seyfarth and Cheney (1993) started me thinking about classifying animal signals as "windows" into the functioning of the brain. You may understand this classification better by taking the last categories first and working backward, because the early signals *lack* features of the later ones.

1. Nonsocial Signals: These signals do not require the presence of other conspecifics: indeed, when other conspecifics are present, they are typically not produced, and If another conspecific approaches when they are being produced, the signaler changes to other signals such as aggressive or courtship signals. Included are probably all of the calling and pair-forming signals of insects and anurans (signals I have called "rapprochement signals") and some territorial and alarm signals. Examples are calling of crickets and katydids and cicadas, flashing of fireflies, and maybe cackling of hens, whistle-snorts of horses, crowing of roosters, and some bird songs. These signals do not involve any of the requirements given below for other kinds of signals, and may not require the kind of learning that involves sensing of the signals themselves.

2. Social Signals: These signals function in interactions with other conspecifics, and I divide them into three categories with respect to what must be going on in the brain.

a. Nonmemory Signals: These signals are produced in the presence of other conspecifics, but they do not involve any memory specific to interactions with the particular conspecifics with which the interactions are occurring (although they may have an ontogeny requiring some interactions with conspecifics that does not involve remembering anything about specific individuals). Included are the courtship and aggressive signals of insects and some other animals such as fish and anurans, signals produced in cicada and other insect leks, and whatever signals might be produced in strictly selfish herds (that is, aside from interactions between relatives or mates within otherwise selfish herds).

b. Memory-Influenced Signals: These signals are produced in the presence of conspecific individuals that are known to the signaller as individuals from past interactions, but no theory of mind is involved (see below). If one accepts, as I do, that no good evidence exists for evolved nepotism except as a result of social learning of the identity of relatives, then these signals represent the first ones in the list that can involve evolved discriminative nepotism (i.e.,

directed at specific individuals). I think that a large number of social birds and mammals exchange these kinds of sign exchange signals under this kind of circumstance).

c. Theory-of-Mind-Influenced Signals: These signals can be influenced by memories of individuals, and in addition by the fact that the interacting individuals have some kind of "theory of the mind" with regard to their interactants, as we humans do. In other words, we have memory not only of attributes of other individuals that enable us to identify them individually, but we include in such memory one or more of the following: (1) memory about the individual histories of different interactants, (2) memory about what immediate circumstances may cause with respect to the reactions of our interactants to particular signals, (3) a kind of general memory about what particular attitudes or motivations may exist in our interactants, and (4) abilities to relate to our own attitudes or motivations to those likely in our interactants.

I don't know where this phrase "theory of mind" comes from, though Seyfarth and Cheney refer to it as "what the psychologists call a theory of mind" and Humphrey discusses it -- not precisely as I have -- especially in his book on consciousness (Rachel told me of a paper by David Premack in the mid-1980's titled "Do chimpanzees have a theory of mind?"). A number of studies have been so directed as to reveal such an attribute if it exists, such as studies showing that chimpanzees ally themselves with an investigator who only accidentally drops food before getting it to them as compared to one who deliberately drops it, and also with another chimpanzee who has had opportunity to see where food is as opposed to one who has not had such an opportunity (and various studies on deception). In studying alarm signals of vervet monkeys Seyfarth and Cheney concluded they did not have a theory of mind. One immediately thinks of Gallup's work with what he called self-consciousness in apes and baboons as possibly relevant.

It is obviously very important to discover which animals signal in which of these ways in connection with which of their life activities. Maybe it's not a bad way to think about communication.

I am inclined at this point to start thinking about listing models of social behavior by animal kind or group, to use in examining any particular animal of interest. For example, I suggested to Andy Richards that he think in terms of the following list of models to see where dolphins are most likely to fit (or most nearly fit), given the necessarily very Incomplete kinds of information that can be gathered about them (e.g., that females sometimes forage alone with or without an offspring, and sometimes in groups; that males go around in pairs or trios and sequester females as such; that larger groups including both sexes also sometimes occur; that particular individuals are seen together a lot, but sometimes any of those same individuals may be associating with other individuals; and there are indications that the kinds of interactions going on involve individual recognition). This description is probably unique to dolphins just because of the extreme consistency of male units of two and three. It is unusual among all animals solely because females with and without offspring may operate either in a social group or

alone. In this respect it is similar to only a very few social mammals and birds, such as—I think—lions, chimpanzees, and some hyenas (?).

One can start thinking of "models", meaning thinking about how different kinds of animals interact socially, expecting that most of them are going to fit 2b above. Some, such as chimpanzees – and probably gorillas and orangutans (and some in this seminar would like to think dolphins as well)—could instead fit 2c above. I note that among these different kinds of animals (models) only wolves, hunting dogs, hyenas, lions, dolphins, chimpanzees, and baboons can be described as resembling humans in being both predators and prey-- and what preys upon lions?

Gibbon Model (male and female apart, territorial)
Horse-Gorilla-Prairie Dog-Hanuman Langur Model (matriarchal groups of related females with
        single males that change from time to time)
Plains Zebra-Gelada and Hamadryas Baboon Model (associated groups of horse-gorilla-
        etc.types of groups that combine when predators attack)
Elephant-Lion-Belding's Ground Squirrel Model (matriarchal groups of related females without
        adult males except during breeding season)
Canada Goose-Wolf-Parrot Model (pairs and offspring separated during the breeding season,
        large groups of pairs and families outside the breeding season)
Savannah Baboon Model (groups of related females with multiple males). This model can be
        expanded by including a lot of selfish herds such as caribou, bison, gnu, wildebeest, etc.
        But it would then need some qualifying because in some of these groups only parents
        and offspring know each other, and then only for a while, and in others there are
        lifetime associations between close relatives—and in most we don't know…
Chimpanzee Model (groups of related males defending territories inhabited by groups of
        unrelated females)

I have undoubtedly made errors or reduced precision by lumping the particular groups I have, and I am stopping without trying to produce an exhaustive list here because I am too ignorant to go on. But you see what I am trying to provoke.

It is possible to continue modeling to include different kinds of "human models" in the way I did in *Darwinism and Human Affairs*, discussing the different kinds of human groupings (see *DHA* figures on pp. 250, 260), going from hunter-gatherer bands through villages, chiefdoms, and finally to nation-states. That in turn starts me thinking about Dunbar's paper that we read earlier, because I want to know not only how many different individual humans we human individuals interact with, but how many we interact with in which of the different ways I have described above. For example, how many different individuals do we know as individuals from what kinds of past interactions? How many only indirectly? How many do we know as potential recipients of nepotistic beneficence? How many in all the different societies that humans have existed in? Do we always treat all human interactants in the same way, with regard to using our "theory of mind"? In all kinds of behavioral interactions? How does our behavior compare with

that of other species in regard to complexity and multiplicity in regard to such interactions? I am sure these thoughts can eventually be taken much further.

Now here is the second set of things I thought about on the trip, which I believe can aptly be appended here. It involves trying to explain how religion (and morality) might have come about during human evolution. I usually choose an unresolved topic to mull over on long airplane flights, or trips where I will have some thinking time, and I chose religion on this trip because explaining all aspects of it has remained a puzzle in my mind. Here are the thoughts I had, which, if Incomplete or wrong, at least seem to me better than any I have had before.

How Religion Originated: Outline of a Sequence or Hypothesis

I. Consciousness generated, with respect to proximate mechanisms of success and failure in (1) garnering resources and (2) actually reproducing. Of course, I see consciousness as generating in the context of developing and using a "theory of mind" regarding social interactants. One could consider precisely why this should have happened, but that has been discussed elsewhere (even though Dunbar says it hasn't), and I'll leave it alone here except to note its relationship to the models of human sociality mentioned above (and see below).

In the ways most likely to be rendered conscious by selection, the above proximate mechanisms involve other humans (the hypothesis of the human brain evolving as a social tool). Conspecifics within groups can aid us (allies, friends, relatives) or thwart us (adversaries, competitors, enemies). Those in other groups never aid(ed) us, except sometimes in getting mates, so throughout most of human history have been enemies only (people in some nontechnological societies that exchange marriage partners between groups even remark that they get their mates from their enemies).

II. Consciousness arose about death of (1) others (relatives, friends, enemies) and (2) ourselves. Consciousness of death in others may have been favored by selection because thereby we can better prevent it. I can't decide if consciousness about our own inevitable deaths has much likelihood of helping us avoid it or behave adaptively with respect to it; if not, it surely could have arisen incidentally because of the value of knowing about it in others. I don't have much of a problem with the notion that consciousness about death might have arisen incidental to the evolution of consciousness in other contexts.

Friends can save us and those of value to us from injury and death; enemies cause injury and death. Even enemies within groups become friends against enemies in other groups.

From all of this arises morality as within-group rules, and as well group-against-group interactions are promoted.

Still left are unknown causes of observable events and phenomena, and consciousness that some causes are unknown. Good from unknown causes implies unknown friends; evil, unknown enemies: God and the Devil; a Supernatural (as cause). A supernatural friend can help

individuals as well as groups, hence merits cultivation; supernatural enemies likewise merit appeasement or a call for help from supernatural friends.

Once all of this has come about, people can be induced to worship God (as friendly unknown cause) and to rely on him to thwart the Devil (unfriendly unknown cause) with reference to personal welfare and death. The transfer can be made to morality as a group phenomenon tied to xenophobia and intergroup strife by going through these following steps:

(1) "Be good or God will punish you" (or not help you against the Devil).

(2) Being good means following rules within the group (tempering the seeking of the proximate mechanisms of success in resource garnering and reproduction).

(3) Being good includes protecting the group against known enemies (members of other groups).

In this one can almost read the British saying (evidently from the time of the Roman invasion there, as my friends in the romance languages say they are unable to trace it to a proper Latin or Greek origin): *Vox populi, vox Dei*, which translates as: "The voice of the people is the voice of God." The implication is that one's collection of social associates -- the continuing society -- decides what is right and wrong. I don't remember ever finding a situation in which this definition did not make sense.

As this stage is being achieved, individuals or subgroups within groups can start claiming to be agents of supernatural causes, and as groups become of the size and sort that benefit from leaders, divine and leader can become correlated, and leaders, especially "divine" ones, acquire power and can use their power and their (however gained) knowledge and cleverness about success to limit the success of others within their group explicitly so as to serve their own interests -- partly by creating the impression of being part of unknown causation and by demonstrating value in defeating other-group enemies and aiding against both within-group and between-group enemies (and eventually by deceiving about both as well); i.e., the relationship between organized religion and politics and government.

If the unknown cause acquires immortality (does not go away or seems to outlast human lifetimes) it can a source of promises somehow to suspend individual demise, therefore a way for individuals to conquer death by appeasement of the unknown cause.

Explaining the Mother Teresas of the world (i.e., saints, "true altruism") can take this form:

(1) Personal altruism begins with nepotism and extends to the beneficence of reciprocity (both direct and indirect), and to commitments and conscience (see RDA unpublished essay on morality).

(2) Altruism (that is, beneficence to others, and commitment and conscience) is encouraged and praised, therefore assumed, affected, and paraded; egoism is despised and accordingly denied.

(3) It is seen as undesirable (or even immoral) to expose acts of actual altruism as selfishly motivated mistakes. The reason is that such a stance will reduce the likelihood of people being altruistic, presumably to the benefit of those who see exposing it as immoral. For the obvious reason, then, sociobiology is immoral...

All of this from the evolution of a theory of mind ...

New References

Anonymous. 1993. Can animals think? *Time* Magazine cover story, March 22, 1993, pp. 55-61.
Radner, D. and M. Radner, 1989. *Animal Consciousness*. Buffalo, New York: Prometheus Books.
Seyfath, R. M. and D. L. Cheney. 1992. Meaning and mind In monkeys. *Sci. Amer.* (Dec), pp. 78-84.

Edited by D. Lahti, 9/5/2021