

**Ray Jackendoff**  
**Language, Culture, Consciousness: Essays on Mental Structure**  
draft of spring 2003  
comments solicited

## **Chapter 4**

### **An Agenda for a Theory of Social Cognition**

Over the past two decades I've had the ambition of working out an approach to social cognition from the perspective of cognitive neuroscience, and I have managed to write a couple of pieces on the problems and the prospects.<sup>1</sup> But now a community of researchers interested in this topic seems to be coming together from many different directions. So it seems worth attempting again to state an agenda for this emerging area of inquiry. I'm going to try to lay out the issues as broadly as I can. I'd like to think about some of the big questions a theory of social cognition could address – especially questions that concern not just scientists but also the general public.

#### **4.1. Social cognition as a cognitive capacity**

What is social cognition from the perspective of cognitive neuroscience? It is the domain of cognitive neuroscience that pertains to a particular sort of subject matter in an organism's repertoire, namely social interaction. For a rough definition, I'll take social interaction to mean the organism's ability to deal with conspecifics and its grasp of interactions among other conspecifics, in the context of larger social institutions. This definition can be applied not just to humans but to animal societies as well, all the way from chimps to ants.

There immediately arises the question of why we should want at all to separate off social cognition from other, more general cognitive processes. A first answer is that in the end, it indeed might turn out not of interest to do so, but we'll never know unless we try. A more constructive answer is that other sorts of knowledge such as language, number, “naive physics”, and “naive biology” have been profitably studied as specialized systems of mind; on the face of it social cognition presents itself as another such content domain. A third answer might point out that many subareas of social cognition have already been examined in terms of specialized capacities: sexual selection, face recognition, morality, Theory of Mind, cheater detection, and so on. So to consider social cognition as a whole is actually a move toward unification rather than one of separation. Finally, treating a mental domain as separate does not mean it is totally isolated from the rest of the mind. Certainly, language would be useless in isolation from thought, perception, and action; and a sense of number would be useless without a conception of things to count. Similarly, brains and hearts are deeply interdependent, but that does not stop us from recognizing

---

<sup>1</sup>Jackendoff 1992, chapter 4; Jackendoff 1994, chapter 15. The present chapter borrows a few passages from each of these. Two more technical papers on intending and on obligation appear here in revised form (chapters 6 and 8 respectively).

that they are separate organs. So if social cognition is to be described as a separate element of mind, by all means part of the description has to be its interfaces with other capacities.

Another immediate impulse against such an undertaking comes from another direction. Isn't it the case that our interactions with others are all determined by culture? And if so, why should we have to worry about a cognitive capacity at all? The reply is that in order for an organism to interact with others, it has to have a mind/brain. Rocks and trees don't have social interactions. And fish and cats don't have the same kinds of social interactions we do. So, even if you insist that our social interactions are determined by culture, it's still important to ask: What is it about the human mind/brain that makes it possible to be influenced and shaped by human culture? (And what is it about the cat brain that makes this *impossible*, despite lengthy exposure to human culture?) It is also worth reminding such skeptics that, although many people still think that *language* is simply a cultural artifact, the last forty years have shown tremendous dividends from studying what it is about the human mind/brain that makes it possible to master language. Here I'm basically going out on a limb, exploring whether a similar approach to society and culture might also bear fruit.

In order for an organism to make use of its capacity for social cognition, of course it has to be able to perceive the environment and act in it. I take it, though, that a theory of social cognition can abstract away from most of the problems of basic perception and motor control and can concentrate on phenomena that are more strictly of social significance. Social significance does penetrate into the perceptual systems when we study phenomena such as face perception, the tracking of eye gaze, and recognition of affect through facial expression, gesture, and posture: all of these fall under "person perception." Motor control gets involved when we study the production of signals such as facial expressions and communicative calls. These are fascinating phenomena in their own right, and their study goes back to Darwin. For me, though, the more important question for a theory of social cognition is what deeper purposes these perceptual and motor phenomena serve. Why should the organism care so much about who it's seeing, and what sorts of interactions to expect from that individual? To answer these questions, we need to think about what *nonperceivable* systems of mental organization make face and affect recognition and the production of signals crucial to the organism.

Social organization shows considerable variation across species, but within most species, it is pretty much fixed across different populations (or so I gather). Within each primate species, especially apes, there appears to be a certain amount of variation from one population to the next. But humans, by contrast, show a vast range of variation in social behavior and social organization among populations, the differences being what we call "culture." Given these differences, human children must learn appropriate patterns of social interaction on the basis of cues provided in the environment by others, whether these cues are provided consciously or unconsciously.

This is where I come in as a linguist. The problem of how a child acquires social/cultural competence bears a deep analogy to the problem of language acquisition, which, as observed in chapter 3, has formed the foundation of contemporary linguistic theory for the last forty years. It

is worth my reviewing this problem before translating it into terms appropriate for social cognition.<sup>2</sup>

Humans manage to create and understand an unlimited number of utterances of their language, most of which they have never heard before. This ability must therefore involve a combinatorial system of principles (or a grammar) in the language user's mind/brain, allowing linguistic structures to be built up from some finite stock of learned elements stored in memory. The grammar is not available to the consciousness of the language user; only its output is available.

The child must acquire this system in the course of learning to speak. The child has no direct evidence for the grammar; again, only its output is available, in the speech of those with whom the child interacts. Learning therefore must involve the active creation of organization in the mind/brain of the learner; it may or may not involve active teaching on the part of those with whom the learner interacts. In order to use speech in the environment as evidence for the grammar, the child must bring to bear inner resources of the mind/brain. Since these inner resources are by definition not learned, they must be a consequence of the inherent structure of human brains, determined by the interaction of the genome with the processes of biological development. Some of these inner resources may be cognitive specializations for language in particular; others may be applicable for purposes more general than just learning language. In principle it should be possible to sort these out. So goes the argument.

Turning to social cognition, we can state an almost parallel suite of issues. The *answers* may or may not turn out to be parallel, but the questions are surely legitimate. The basic

---

<sup>2</sup>It is not so new to draw a parallel between language and social interaction. John Rawls (1971) pointed out that we can make intuitive judgments of the justice or fairness of a situation, often in instances where we cannot explicitly state the principles behind our judgment. He drew an analogy to our ability to make judgments of the grammaticality of sentences, citing Chomsky's then-new theory of grammar. This parallel has been picked up more recently by John Mikhail (2000), who has been exploring a theory of morality modeled on linguistic theory.

I think Rawls was on the right track. But, like many people in many different fields at the time, he drew too shallow a parallel between linguistics and his own concerns. An ability to make grammaticality judgments is not the basic point of linguistic competence; it is a side effect of being able to use language for communication. By contrast, one's ability to make judgments of justice or morality *is* the point of moral competence: it helps determine how one behaves toward the person being judged and how one guides one's own actions. Here I wish to take a deeper starting point for the analogy, going back to the first principles that motivate the theory of generative grammar. A closer antecedent for my approach is Macnamara 1991, who draws a parallel between moral reasoning and intuitive geometry.

observation is that humans manage to participate in and understand an unlimited number of social interactions, most of which they have never encountered before in the exact same form. (Among these, of course, are linguistic interactions, so the language system has to embed in the larger social system.) The ability to interact socially must therefore involve a combinatorial system of principles in each individual's mind/brain, which makes it possible to build up understanding of particular situations from some finite stock of stored elements. There are certainly some normative principles of social interaction that individuals can state explicitly; here social cognition differs from language. But there are also likely to be principles of interaction and social understanding that, like the f-rules of language (to use the term of chapter 3), are quite inaccessible to consciousness (i.e. "intuitive").

The child must acquire the system of principles in the course of being socialized. The child has only partial evidence for the system; for the most part only its output, actual social behavior, is available to the child. Learning must therefore involve the active creation of organization in the mind/brain of the learner; for social interaction, unlike language, there may well be considerable explicit teaching on the part of those with whom the learner interacts. In order for the child to use tacit understanding of social behavior in the environment as evidence for the system of principles, the child must bring to bear inner resources. Since these inner resources are by definition not learned, they must be a consequence of the inherent structure of human brains, determined by the interaction of the genome with the processes of biological development. Some of these inner resources may be a cognitive specialization for social interaction; some may be applicable to purposes other than learning a social system. In principle it should be possible to sort these out. Table 4.1 summarizes the parallels.

Table 4.1. Parallels between language and social cognition

- |  |   |
|--|---|
| • Unlimited number of understandable sentences   | • Unlimited number of understandable social situations  |
| • Requires combinatorial rule system in mind of language user  | • Requires combinatorial rule system in mind of social agent  |
| • Rule system not available to consciousness   | • Rule system only partly available to consciousness  |
| • Rule system must be acquired by child with only imperfect evidence in environment, virtually no teaching | • Rule system must be acquired by child with only imperfect evidence, only partially taught         |
| • Learning thus requires inner unlearned resources, perhaps partly specific to language                    | • Learning thus requires inner unlearned resources, perhaps partly specific to social cognition     |
| • Inner resources must be determined by genome interacting with processes of biological development        | • Inner resources must be determined by genome interacting with processes of biological development |

The value of the parallel, of course, depends on how complex the social behavior of humans proves to be: whether its degree of complexity is comparable to that of language, or whether, as is often thought, it consists of relatively simple patterns. If the latter is the case, then the opening premise of the argument, “unlimited number of understandable social situations,” is false, and the rest of the points reduce to relative triviality. However, to the degree that social behavior is complex and subtle, the parallel goes through and the rest of the issues are indeed of interest.

As an incentive toward accepting this opening premise, readers should bear in mind that many people still think of language as simple too, in ignorance (or irrational rejection) of the sorts of complexity represented in Figure 1.1 for a trivially simple sentence (see sections 1.2 and 1.3). Moreover, I suspect that those who are skeptical about the indefinite variability of social interactions are not looking at a fine enough grain of interaction. For instance, section 5.2 examines the action of shaking hands, a social interaction that seems on the face of it totally stereotyped. We will see that even here there is immense complexity and subtlety. I cannot help but think the complexity scales up phenomenally when dealing with real social life. Traditions of “thick” description in anthropology (ref?) and sociology (Goffman 1974) seem to me to reinforce this point.

If we follow this line of inquiry in parallel with language, all aspects of cognitive neuroscience come to bear on the problem of social cognition. In addition to studying the characteristics of social/cultural behavior and inquiring into universals of human culture (the counterpart of studying grammars of languages), we can ask about the neurological and genetic bases of social cognition (the counterpart of neurolinguistics), and about the cognitive processes of social cognition, that is, how the brain processes, accesses, and stores social information in real time (the counterpart of psycholinguistics). We can ask about the course of the child's social development (the counterpart of developmental psycholinguistics); and about the consequences of various brain deficits for social cognition – whether genetic deficits (perhaps autism) or deficits acquired through injury (e.g. Phineas Gage). In short, all the angles available for studying the language capacity have analogues in the capacity for social interaction.

One additional line of evidence is available for human social cognition that is not available for language: comparative ethology. Primate societies are highly structured and vary along various dimensions from species to species. This hints at a strong innate genetic basis to their social organization. Moreover, much of primate social behavior looks quite familiar to us, involving issues of kinship, dominance, alliance, group membership, and reciprocity (Goodall 1971, Cheney and Seyfarth 1990, Byrne and Whiten 1988, de Waal 1996, and many others). This suggests (following Darwin) that behind human culture lies a firm foundation of primate evolutionary ancestry. So by standard comparative methods we can form hypotheses about the ancestral great ape social repertoire, and we can ask what tricks evolution had to add to the ancestral repertoire to get modern chimps, bonobos, gorillas – and us. Thus the issue for the evolution of social cognition is not just what problems early hominids had to face (as stressed by

Tooby and Cosmides 1992 for instance), but also what earlier ancestral primates (and mammals before them) faced as well, and therefore on what prior solutions hominids were able to build.<sup>3</sup>

A further dimension has not played much of a role in the theory of language, but it could – again with a parallel in the theory of social cognition. It may be that some linguistic and social phenomena are not strictly speaking part of the abilities of the individuals taking part in them, but are rather just an emergent property of individual behaviors. For instance, I'm pretty sure that cooperative behavior among ants is not the product of overt agreement among them or some fancy theory of mind. Rather, it's just the product of a number of cheap tricks – automatic responses to particular actions by conspecifics such as emitting smells. I don't care whether we want to call this collective behavior social cognition or not, but it's of interest to explain how the social dynamics of ant colonies arises through the interaction of relatively stupid individual agents who are not gauging the overall consequence of their actions. Now it might be that some aspects of human cultures have this character too. For example the behavior of the stock market may be like this – usually its gyrations are not the direct effect of planning, but rather are emergent from the behavior of many independent agents.<sup>4</sup> Most likely language change (such as the changes between Shakespearean and modern English) is also attributable to such effects. The point is that we should be careful not to overinterpret human social behavior.

Still, we shouldn't underinterpret it either. The example of linguistics again comes to mind. Linguistic behavior, at all levels from phonetics to pragmatics, has proven far more complex than anyone would have expected 50 years ago. It's just that most of the complexity of language falls below the radar – it is transparent to us; and it is only with training in linguistics that we begin to notice it. As outlined above, I suspect the same is true of social behavior. The trick is to figure out *exactly* what capability should be ascribed to the individual, and to show how that results in the observed group dynamics. The part due to group dynamics corresponds to the traditional sense of “social construction” of knowledge, whether of language or of culture; it emerges out of the collective practices of the group, without anyone necessarily intending it to do so.

This cognitive, biological, and evolutionary approach is not the predominant way that culture is studied. At least in America, anthropology and sociology are dominated by the view that humans are totally a product of their culture, and that it is meaningless to claim that culture is the way it is in part because of human cognitive abilities. Ehrenreich and McIntosh 1995 and

---

<sup>3</sup>Although I think evolutionary considerations are important, I don't think every aspect of human social organization can be explained, much less discovered, by appeal to first principles of evolutionary psychology. Little of the detailed structure of language can be explained on evolutionary grounds; indeed, if we insisted on starting from the point of view of evolution we would be hard pressed to discover much of it.

<sup>4</sup>Merlin Donald (1998) claims that a great deal of human social behavior, including language, is of this character. I'll grant him some.

Pinker 2002 document how widespread and influential such attitudes are, not only among scientists but in politics and ordinary life.

There were originally good reasons for such attitudes. Degler 1991 documents the pervasive assumption in the late 19th and early 20th century (held even by Darwin and Humboldt) that races and ethnicities are sharply distinguished in intelligence and moral capacities – with northern Europeans naturally at the peak of both. Such views comported with the rampant colonialism of the period, and were also invoked to reinforce anti-immigration legislation (unless of course the immigrants were from the right place). Degler relates how the anthropologist/linguist Franz Boas fought fiercely for cultural relativism in the interests of resisting such racism; the well-known work of his students Margaret Mead and Ruth Benedict was intended to demonstrate the complete arbitrariness of cultural institutions and moral systems. But it is one thing to assert that a culture’s technological development and military prowess are a consequence of cultural differences and ecological opportunities rather than differences in inherent intelligence (an argument taken up again in our time by Jared Diamond (1997) against recent resurgences such as Herrnstein and Murray 1994) – and it is something else altogether to conclude that *no* aspects of human cultural capacity have an inherited basis, and that there is no inherent human nature of relevance to the social sciences.<sup>5</sup>

It is still worse to conclude, as I gather many in anthropology have done, that it is impossible – and wrong – to attempt to go beyond a radically local, relativistic, and contextualized perspective on the culture one is studying (these attitudes are documented by Brown 1991, Zuriff 1998, Ehrenreich and McIntosh 1997). This renders impossible any sort of scientific crosscultural comparison. Some rejoice at this. I don’t. It is indeed objectionable to take a stance of scientific paternalism and condescension towards other cultures – “we enlightened scientists know better, and aren’t you exotic.” This only mirrors the colonialist and imperialist attitudes of a century ago. But as with the description of other people’s languages, I see no problem with doing the best, most honest job of description we can, respecting native intuitions and recognizing our fallibility. That’s just good science. Moreover, attempts at such cross-cultural comparison such as Brown 1991 and the monumental Eibl-Eibesfeldt 1989 reveal a rich tapestry of universal or at least widespread patterns of social behavior and organization.

In short, anthropological description and interpretation does have much to offer us in terms of data about social behavior and social organization. But a theory of social cognition

---

<sup>5</sup>A recurrent mistake, one so common that it’s tempting to attribute it to human nature, is to draw a parallel between animal species and human races. Even E. O. Wilson, a pioneer of the cognitive and evolutionary approach to social organization, succumbs to this in his *Sociobiology* (1975). The whole book is about innate species-specific differences that govern social behavior. Then suddenly in the final chapter on humans there are passages about heritability of racial and cultural differences. Here I am interested in the characteristics of the species as a whole (as is Wilson, most of the time).

becomes quite different when you take seriously the cognitive capacity that lies behind the overt phenomena that anthropologists study.

Perhaps betraying my training as a linguist, I'm most interested in social cognition from a formal or functional point of view. At our present stage of understanding the brain, we may be able to localize some social function, say face recognition in the right parietal lobe or moral judgment in prefrontal areas; or we may be able to find a neurotransmitter that enhances aggression or affiliative behavior, and pinpoint its locus of action in the brain. But figuring out how the whole system works is probably better pursued at a level of abstraction somewhat distant from the neurons. Again I see this as analogous to the study of language. As I have emphasized several times in the previous chapters, we know a fair amount about localization of different aspects of language function in the brain, but for the moment I think it's fair to say that it will be a long time before we understand how the details of Bulgarian case marking and Mandarin tone in language – or the notion of ownership in social cognition – are instantiated in the brain. For the moment it seems more fruitful to look at these problems from the perspective of an abstract structural grammar or internal logic, and to put off issues of neural instantiation for a while. This doesn't preclude applying the tools of neuroscience by any means – there's lots to be learned there as well. But the two efforts ought to run in parallel.

#### **4.2. The physical and the social/personal domains**

Where is social cognition to be localized in the functional ecology of the mind? My basic hypothesis is that social cognition is one of the central systems of cognition: it is a major department of the level of conceptual structure, standing alongside and interacting with the understanding of physical space. The latter involves concepts of physical objects which are located in three-dimensional isotropic space, which move in this space, and which exert forces on each other. Among the physical objects are natural objects like rocks and trees and rivers, functional objects with affordances for use like bicycles and tables, and animate objects like ants and worms and rats and tigers. The animates, unlike the rest, are conceptualized as capable of unpredictable self-initiated motion (i.e. volition) – and therefore, perhaps of desires and intentions as well.

The social domain at its purest consists of *persons* – individuals with whom we can have social relations. It encodes the relations and actions among them *as* persons. Like all concepts, the concept of person has a certain amount of leakage at the boundaries. Pets probably count as (sort of) persons, as do “personified” animals in Aesop’s fables; but the mosquito that’s buzzing in your ear, though animate, certainly doesn’t count. In the other direction, an all-too-common social tactic for endorsing ruthless behavior toward another group is to characterize them as animals, i.e. not persons, and hence not qualifying for social relations.

The social domain, in contrast with the physical domain, is not an isotropic space. In physical space, between any two objects that are not touching, there is an intermediate region of

space that can be occupied by another object. Such a notion makes no sense in social space – there are no “intermediate spaces” between people. To the extent that there is a notion of distance, it is “social distance”, measured in terms of divisions of kinship, class, status, group membership, and degree of intimacy or alliance. Of course there are notions of force, coercion, and constraint in the social domain, sometimes connected with physical force or the threat thereof, sometimes not.

*People* are conceptualized as occupying both the physical and social domains – the duality signified by the traditional division between body and soul. One way to demonstrate this duality of conceptualization is to notice the culturally widespread belief in entities such as spirits, ghosts, gods, and souls that survive death. These are beings who lack definite physical bodies, yet have social relations with people and with each other. Hence they exist in the social domain, but not the physical. In addition, we have no problem conceptualizing persons coming to inhabit different bodies through reincarnation, metamorphosis, or body-switching. In each case, personal identity – in the social domain – is thought of as being preserved despite radical physical change.<sup>6</sup> A further argument for duality and separation of the two domains is the extreme difficulty people have in accepting and reasoning within a materialist philosophy of mind – thinking about persons as defined only in physical terms. Similarly, people show extreme discomfort with notions like golems, humanoid computers, and the like – physical objects which suddenly sprout social identity or personhood. That is, conceptually there is a transcendental difference between the physical and the personal, one that is virtually impossible to erase.

Actions on the social plane are on their own unobservable. They become observable only through their linkages with the physical plane.<sup>7</sup> Some physical actions such as eating and walking make sense on their own, but some, such as performing religious ceremonies and shaking hands (chapter 5), make sense only as instantiations of (or symbols of) actions on the social plane.<sup>8</sup> Choices of costume or speech style can be used to signify social roles, but they don’t constitute the roles they symbolize – even if people sometimes act as if they do (e.g. the military uniform or the flag standing for the country). Even social relations with disembodied spirits involve physical actions; we call them magic or ritual.

It is not as though people consciously separate these two planes. Rather the issue is the functional decomposition of social concepts, how to “carve a natural joint” in our intuitive modes

---

<sup>6</sup>Notice that such conceptualization depends on there being two independent domains – what would it mean to say of two merely physical objects: “This fork and this spoon just exchanged identities with each other”?

<sup>7</sup>This corresponds to Searle’s (1995) notion of the logical priority of “brute facts” over “institutional facts”

<sup>8</sup>This corresponds to Searle’s (1995) notion of a physical action “counting as” an instance of a social action.

of thought. In fact, the idea of entities and actions being simultaneously formulated and interpreted in two parallel linked planes – below the level of awareness – is not so unfamiliar in the theory of cognition. A commonplace example comes from Pustejovsky 1995. Compare a brick and a book. At the physical level, they have similar properties: you can lift them, move them around, stack them up, and so on. But a book has a whole other set of properties that comes from the fact that it is conceptualized as containing *information*: you can read it, copy it, analyze it, understand it, and so on. Information, in order to be conveyed, requires a physical instantiation. But some instantiations are not objects, for instance a linguistic utterance or a gesture; these too can be perceived, copied, analyzed, understood. Thus it makes sense analytically to think of a book as partaking of two different natures in parallel, each of which can exist without the other. And actions involving a book, such as reading, take place on both planes at once: one is both moving one's eyes over the physical page and taking in information.

A deeper (but not unrelated) example is language, which proceeds in two parallel planes, phonology and meaning. One can associate different phonology with the same meaning by switching languages; one can have sound dissociated from meaning (nonlinguistic sounds and nonsense syllables); and one can have meaning dissociated from sound, i.e. thought. Meaning (at least others' meaning) is unobservable without being linked to speech. Yet this parallel organization is largely transparent to experience: it is common to think of meaning as just an inherent part of the spoken word. So it is, I think, with persons and their bodies. And just as one can mislead about one's thought by uttering sounds linked with some other thought (i.e. lying), one can mislead about one's social intentions by producing physical signals consistent with other intentions, say through facial expression, gesture, or disguise.

I think it is important to separate the notion of the social domain from the Theory of Mind. Stereotypically, in human social relations, we attribute a mental life to the persons with whom we interact. But it is possible for us also to attribute a mental life to animate entities with which we have no social relation, say when we are watching a tiger stalking an antelope. We do not hesitate to say the tiger has beliefs and intentions (whether we are correct or not is beside the point: we are using *our* theory of mind here). Conversely, many aspects of social relations do not require a theory of mind: for a person to be a member of a certain clan and therefore to have certain obligations, it does not matter what we think they believe or desire, it is just an objective social fact. Moreover, it makes sense to attribute some sort of social cognition to monkeys, who, according to much current thinking, lack theory of mind. So although theory of mind clearly plays an important role in human social cognition, these two aspects of cognition are not coextensive.

### **4.3. Cooperation**

One of the major tasks for a theory of social cognition and action is to delineate the elements and relations in the social domain and their potential linkages to physical action. I want to spend the rest of this chapter exploring some of these. Social psychology has dealt a great deal

with person perception; sociobiology and evolutionary psychology have rehearsed at length issues like sexual selection and altruism. I'd like to take up some other issues that are a bit more distant from direct evolutionary justification.

One fundamental opportunity presented by interaction with conspecifics is cooperation. In my work on lexical semantics over the years, I've studied a lot of words that express intention and social coercion, for instance *intend* and *persuade*. In particular, the concept expressed by 'X intends to do Y' can be glossed roughly as 'X has a commitment to performing action Y' (see chapter 6). But I was stumped by other words like *cooperate* and *agree*, until I came across John Searle's (1995) proposal that we are capable of conceptualizing what he calls 'joint intention': not just '*I* intend to do such-and-such,' but '*we* intend to do such-and-such, and my role in it is such-and-such.' Perhaps the simplest case for imagining this is moving furniture together: my lifting one end of the couch and pushing makes no sense outside the context of your lifting the other end and pulling, so that the jointly intended action of moving the couch out the door takes place.<sup>9</sup>

The notion of joint intention seems just right to characterize cooperation and agreement. It is a necessary component of any sort of transaction, trade, or contract. And contracts of course include marriage contracts, which are at base (I suppose) declarations of joint intention to maintain sexual exclusivity (in monogamous societies at least). Clark 1996 argues that joint intention is also a basic aspect of linguistic communication. Using language involves not just a speaker imposing information on a hearer, it's a speaker and a hearer performing a joint task of getting information across.

A joint intention, of course, is a curious phenomenon: it sort of pretends that two people are sharing their minds. Grosz and Sidner 1999 point out that a participant in a joint task does not have to know all the details of the other's action – just enough to insure that the interaction takes place properly. The best approximation individuals can actually make is for each of them to get assurance from the other that the joint intention is shared. So there have to be signals for offering cooperation and for uptake of the offer. These can be most explicit in things like signatures on a contract, a handshake, or at least saying *okay*. Or they may be more subtle, such as little inflections of body language. (Bangerter and Clark 2003 trace in some detail how speakers use expressions like *okay*, *right*, and *uh-huh* to signal that they are in synch.)

The notion of joint intention raises lots of questions. Exactly what sort of cognitive structure is this? What are the signals of offer and uptake, how are they produced, and how are

---

<sup>9</sup>This proposal is not without controversy. For instance, Grosz and Sidner 1999 attempt to eliminate the 'we intend' from a characterization of joint action; on the other hand, as Hobbs 1999 points out, they replace it with the primitive SharedPlan, which has the same effect.

they perceived? Is joint intention necessary for cooperation?<sup>10</sup> When chimps and wolves cooperate, do they need such cognitive structures? (As I said earlier, I don't think ants do.) If chimps and wolves don't conceptualize such a thing as a joint intention, what evolutionary precursors *do* they have that enable them to cooperate to the extent they do?

One important precursor is surely joint *attention*: observing that the other individual is looking where you're looking – which doesn't require theory of mind (see chapter 7). Another place to look for precursors might be sexual behavior, both courting and actual intercourse, both of which call for a certain amount of physical coordination between individuals, though not theory of mind. Still another would be mother-infant interaction.

When we think about the formation and the execution of joint intentions, we're immediately into issues of game theory. Prisoner's dilemma sorts of situations arise from the possibility of defection – abandoning a joint intention and leaving the other participant in the lurch. And joint intentions also open opportunities for deception: letting the other participant continue to think one is pursuing the joint project, while actually planning defection at some later point in time. So theory of mind and so-called cheater detection (Cosmides and Tooby 1992) – psyching out whether the other participant *really* shares the joint intention – come into play quite naturally.<sup>11</sup> Section 4.5 will offer a situation where there may be a more severe disconnect between a conceptualized joint intention and each participant's individual intention.

A sort of opposite polarity to joint intention is competition. Each participant knows the other is out to get him, so there is in some sense a "joint project" of exploiting each other. This is more complex than just "I'll exploit you and protect myself," i.e. plain aggression, because it includes a theory of the competitor's goals. But competition is not symmetrical with cooperation: one cannot take advantage of a competitor by defecting. Usually the only way to opt out of competition is to surrender, unless it is by a joint decision (i.e. a cooperative decision) to abandon the activity.

#### **4.4. Affiliations: Kinship, alliances, dominance**

---

<sup>10</sup>Perhaps an example on the borderline is the situation where you're walking toward someone in the street, and both of you swerve slightly to avoid a collision. I'm not sure this requires a joint intention. But if by chance you both swerve in the same direction at the same time, now there begins this funny little dance, replete with eye contact and facial expressions, until you manage to find nonintersecting trajectories. This has more of the signs of a joint task.

<sup>11</sup>Mother-infant interaction is an interesting case here: The mother interprets the action as joint intention – that's what makes her feel bonded. But I doubt the infant it yet capable of seeing it that way!

Cooperation and competition are relations among individuals that can shift from moment to moment as circumstances dictate and actors undertake different activities. These relationships lie on top of a web of longer-lasting relationships among individuals.

Perhaps the most obvious of these is *kinship*. In every culture, each individual is in a special relationship with her or her parents, children, spouse(s), and siblings. Many aspects of this relationship arise clearly from the mammalian heritage, in which the parents (or mother alone, depending on the species) must take care of the young for some period of time. Evolution has provided us, like other mammals, with patterns of perception and behavior that make this care possible and basically pleasurable.

Kin altruism extends beyond parent-child relationships to include siblings and potentially even more distant related kin. Theoretical models based on the “gene’s-eye view” predict such relations: as kin share genetic material, acts done on behalf of kin can lead to proliferation of one’s own genes to some degree. (The mathematics of such relationships are due to Hamilton 1964; see also discussion in Dawkins 1989.)

In human societies, kinship bonds are extended to more distant relatives as well as immediate family. Many cultures have elaborate customs, obligations, and rights associated with being in particular kin relations. For example, every culture has an incest taboo, but its precise extent varies from culture to culture – with which extended kin sexual relations are forbidden and with which they are permitted or even encouraged.

Although we have perceptual cues for who is in our immediate family (the people we live with), we don’t have any such cue for more distant relatives. We rely on someone *telling* us we’re related, and miraculously we come to feel the bonds of kinship. People easily can feel warmth towards distant cousins they have never heard of, whom they meet for the first time at a family reunion. Consider too the bonds often felt by an adopted child toward a newly discovered biological parent. Such examples show that the bond of kinship is not just *perceptual*, but *conceptual* as well. When we conceive of being kin to someone, we feel and behave differently toward them.

A different sort of relation is that of allies or friends, unrelated individuals between whom, for whatever reason, there is a voluntary and lasting commitment to cooperative activity. The flip side is rivals or enemies, between whom there is a lasting commitment to competition. In both cases, participants know what they can count on from the other. I am not aware of discussion in the literature on crosscultural similarities or differences in such relationships. My impression is that in many cultures they can be formalized by oaths and the like – institutionalized agreements to establish the mutual relationship. Such relationships are documented in the primate literature as well (Goodall, Smuts).

More prominent in the literature is discussion of dominance, a relation between two individuals whereby one the (subordinate one) regularly defers to the other (the dominant one) in

matters of food choice, sexual selection, grooming partners, and so forth. Dominance is often based on size and aggressiveness, but it doesn't have to be. For instance, it can depend on kinship relations: in many species, the children of highly ranked mothers often inherit high rank. This means that dominance cannot be a purely perceptual relation: it too needs a conceptual basis.

In animal societies, dominance relations often fall into a linear order: if A is dominant to B, and B is dominant to C, then A is also dominant to C; and every individual in the group has a distinct place in the "pecking order." Dominance hierarchies characteristically remain stable over time, but subordinate individuals may mount challenges which, if successful, rearrange the pattern. Cheney and Seyfarth (1990) present an extended discussion in which they conclude that the species they study, vervet monkeys, understand the dominance hierarchy as a linear ordering and not just a collection of pairwise relations. Individuals know not just their own relations to each other monkey in the group, but also the relations of other monkeys to each other. (Cheney and Seyfarth also show in several ways that the monkeys appreciate kinship relations among other individuals.)

In human societies, dominance relations are pervasive too. But rather than there being a single pecking order, dominance can be organized along many different dimensions, such as parent to child, teacher to student, boss to worker, ruler to subject, celebrity to fan, and in many cultures, husband to wife. It seems to me that when larger-scale human dominance hierarchies develop, they differ from animal hierarchies in tending to be pyramidal rather than linear, with a top person dominant to a number of relatively equal subordinates, each of whom is dominant to further subordinates, and so on. This drastically expands the size of the group over which dominance can be handled; but the basic notion of a stable asymmetrical relationship based on deference of one individual toward the other bears a strong resemblance to the animal model.

All of these relationships require that you keep track of who is who in your social milieu. Presumably this is the functional motivation for the perceptual specializations in face and voice recognition.

#### **4.5. Groups**

Another kind of lasting affiliation, forming one of the most important elements of social structure, is group membership. The fundamental premise of the logic of groups is that some set of individuals constitutes a group, and everyone else is not a member. What counts as a group? On one extreme are mere aggregations: for example, the people who happen to be on the bus with me at the moment don't constitute a group. Then there are what might be called "affinities", say the baby boomer generation. Again, although baby boomers might share certain common goals, I don't think baby boomers feel any special affinity for other baby boomers such that they would be said to constitute a group in the requisite sense. On the other extreme are formalized groups such as a club, an orchestra, or a religious congregation. Families, extended families, and clans are particular sorts of groups which add kinship relations on top of the basic premise.

The point of groups is that one's actions toward others are conditioned not by who they are as individuals, but whether they are members of the group or not. The most basic principles appear to be the following axioms. They pertain to primate groups, and to every kind of human group from teenage cliques to nations, with professions, religions, and social classes in between.

Axiom 1. Other things being equal, if you are a member of my group, I will behave favorably toward you, e.g. I will be willing to cooperate with you, and I will expect the same from you.

Axiom 2. Other things being equal, if you are not a member of my group, I will behave unfavorably toward you, e.g. I will compete with you, and I will expect the same from you.

Given this logic, it's important to be able to determine who's in and who's out of one's group, especially when groups get so large that members are not necessarily acquainted with everyone else in the group. Members of human groups often make themselves more easily identified by adopting characteristic dress, customs, and manner of speaking.

An interesting hypothesis is that a group is conceptualized as a "super-individual." There seem to me to be two criteria that characterize the entities I want to regard as groups, e.g. clubs but not the aggregation on the subway:

Criterion 1: The group on occasion acts *as a group* or *in the name of the group*, regardless of whether all the members are involved in the action.  
and/or

Criterion 2: The group's existence does not depend on particular people being members; i.e. members can come and go but the group remains in existence as "the same group."

These criteria are what give a group its identity independent of its members.

The view of a group as a "super-individual" helps explain some of the hallmarks of group membership. Just as one has one's own self-esteem as an individual, one has self-esteem that derives from one's group membership – from the "joint self-esteem" of the group. Thus members experience feelings of pride in their own group, a sense of its superiority to other groups, and on occasions a partial loss of individual ego within group identity. Moreover, groups characteristically stage events that reinforce group identity and allegiance. These include rituals that grant membership or status, such as coming of age ceremonies, coronations, marriages, and award ceremonies. But funerals and football games can also play this role in the ecology of groups. (Why oh why does this work? There has to be something in human nature....)

Within this "super-individual", a group member is conceptualized not as an individual, but as an instance of a category. Thus, as in all other cases of categorization by humans, there is a pressure to conceptualize all the instances as being alike – to reduce everyone in the group to an essentialized stereotype. (In case intuition doesn't make this abundantly clear, see Hirschfeld

1996 and McIntosh 2002 for discussion of racial and ethnic essentialism.) This pressure is not confined to one's conceptualization of other groups: within the group there is also a pressure for everyone to be alike.

In the human case (far less so with animals), one typically identifies with numerous overlapping and hierarchical groups. Should I act at any particular moment as an academic, a cognitive scientist, a linguist (one chain of embedded groups), an American, a New Englander, a resident of Belmont, Massachusetts (another chain), a Jew, a conservative Jew of Eastern European descent (yet another), a musician, a member of the orchestra, a member of the wind section, one of the clarinets (still another)? And in terms of which of these groups am I identified by the people I am interacting? As with other categories, there is often a presumption of a sharp distinction between members and nonmembers, a demand of some sort of "purity." (In academia, for instance: I'm not considered exactly a linguist because I think about psychology, but I'm not considered a psychologist because I don't run experiments.) And often groups enforce this purity by establishing systems of admission to the group and procedures for determining descent (including, in academia, intellectual descent). An insistence on group purity combined with the inevitable mixtures of group memberships lies behind the outcast status often accorded to those who make the mistake of being of mixed background.

The view of a group as a "super-individual" also makes it easy to understand the relations among groups. Like an individual, a group can exert dominance over another, compete with another, or form alliances for cooperation with another. In turn, these relations are "inherited" by members of the group. Thus a member of a dominant group will presume personal dominance over a member of a subordinate group (one of the bases of racism and the like). And members of allied groups are more likely to show affiliative behavior than members of competing or hostile groups ("My country is an ally/enemy of your country; therefore you are my friend/enemy").

In order to assure the continued cohesion of a group, it is necessary to enforce Axiom 1 against members who choose not to cooperate. Groups therefore invariably have a code of conduct – a set of normative principles, which may be explicit or implicit. I'll ask about the general character of normative principles in section 4.7. My sense is that the code of conduct is conceptualized as a *joint commitment* of the members (an ontological type I don't recall seeing in the literature, but related to a joint intention). This means that punishment for violating the code is understood as collectively imposed by the group. In smaller, less formal groups, the sanctions are actually imposed collectively – everyone snubs or retaliates against the offender (Ellickson 1991). Larger, more complex groups have to invent institutions that grant authority to certain individuals to impose sanctions on behalf of the group. One of the worst sanctions that can be imposed, universally I believe, is expulsion from the group – the victim "loses his identity."

I should emphasize, though, that although the "will of the group" is conceptualized as a joint commitment, not everyone need be individually committed to it. An extreme case would be a society in which the authorities fraudulently purport to represent the will of the people, and the

laws fraudulently purport to be for the benefit of the group; but everyone plays along with the charade in public out of fear rather than commitment.<sup>12</sup>

In studying other cultures and in engaging in our own, we take for granted all these parts to the logic of groups. This raises the usual developmental issue: do children learn all this? Or do they understand the logic of groups innately and just plug into it any groups with which they come to associate? Given that a parallel though less complex instantiation of this logic appears in primate societies, I would be inclined to vote for a substantial innate component.<sup>13</sup> We might also ask if there are socially impaired individuals who never understand this logic, or if it can be disrupted by brain damage.

What sorts of things might be learned? One important variable in the customs of a group is the degree to which it enforces conformity and sublimation into the group. For instance, it is often said (e.g. Gardner 1983) that American society, at least outwardly, encourages individualism and tolerates nonconformity, whereas Japanese society tends to discourage both. Another variable seems to be the intensity with which Axiom 2 is applied (at least consciously), and to which other groups. For instance, Islam in the Middle Ages and in the Ottoman Empire seems to have had a live-and-let-live attitude toward other religions and ethnicities, contrasting sharply with contemporary fundamentalist Islam. Similarly, the catastrophe of the early 1990s in the former Yugoslavia can be seen in large part as coming from a radical shift in the public face of this parameter, from relative tolerance to intense intolerance. I don't think I need to multiply the horrible examples. Publicly, American society applauds efforts to teach tolerance. But this impulse is far from universal, and even those who advocate tolerance often act otherwise. In any event, the settings of both these variables (in both their overt and tacit manifestations) must be learned by individuals from their culture.

A further issue is the dynamics of groups. So far I have talked as though groups last stably forever. But it is also necessary to understand how groups come to be formed, how they disintegrate, and how they fission. Here is, I suppose, one place where politics enters (another is challenges to the dominance hierarchy).

#### **4.6. Framing**

Structures like cooperation, competition, dominance, and group membership have to be integrated dynamically into one's understanding of the situation from moment to moment, to help

---

<sup>12</sup>A different case is when the rulers are fraudulent but the public does *not* recognize it; here citizens do have the joint commitment, and only the rulers are defecting. Is this a characterization of the US these days?

<sup>13</sup>Famous experiments by Sherif and Sherif (1966) showed how children who were arbitrarily divided into two groups spontaneously developed all the stereotypical symptoms of group identification and group competition.

determine one's course of action. One of the elements of integration might be called *framing* (a term borrowed from Goffman 1974; Minsky 1975 uses the term similarly but with far more limited scope). A typical example of a frame is a concert, in which performers, audience, and ushers have specific roles to play throughout the course of the event – and all participants have to conceptualize it that way in order to understand what is going on and how to behave appropriately in their respective roles. Once the concert is over, the roles are no longer relevant, and the participants drop the frame.

A more general sort of frame is illustrated by games -- which, as far as I know, are found in all cultures. What we see most saliently in a game is the element of competition: two (or more) people knowingly trying to outdo each other. But this competition is set within a larger framework of cooperation: the participants agree to play, and agree to abide by set rules and a presumption of fairness. If a game were *only* competition, poker players would simply be trying to steal each other's money rather than sitting around civilly at tables. The card sharp is of course being deceptive about the frame of cooperation (or placing it in a still larger frame of exploitation): he *is* trying to steal the others' money.

If in the course of a game, someone breaks a rule, the game's progress is suspended till matters are set right. Sometimes the game itself has "metarules" to deal with such cases, but sometimes instead a violation of the rules signifies a defection from the larger frame of cooperation, and degenerates into haggling or even violence.

If the framing of a game is competition within cooperation, can there also be the opposite: cooperation framed inside of competition or hostility? This seems a good characterization of *bargaining*: each participant wants things the other has and is trying to get as much as possible while minimizing his own losses. But this competition is carried on with a facade of civility, so the participants don't end up beating each other up or stealing outright. (This framing corresponds nicely to Alan Fiske's (1991) "Market Pricing" and Jane Jacobs's (1994) "Commercial Syndrome", of which more presently.)

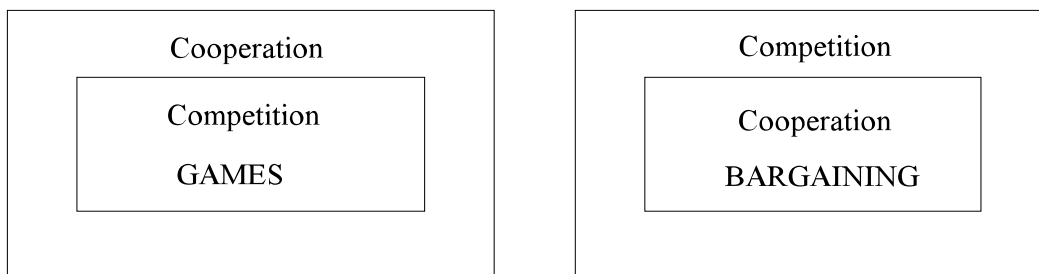


Fig. 4.1. Nesting of frames

We reach another level of complexity in team sports, where each team forms a group whose joint intention is to compete with the other team, all within a frame of cooperation with them. Think about what this entails in real time. In football or hockey, an individual team member at any moment has to gauge his or her actions with respect to multiple individuals, establish joint sub-intentions *very* fast and coordinate them with physical activity. A more highly structured game like baseball adds turn-taking in addition to this constant reframing, where the turns are defined both at the level of individual batters and at the level of teams. The whole complex overarching frame of a team game can be adopted on the spur of the moment, as when kids choose sides in a pickup game. Or it can constitute a lasting relationship, where teams are affiliated with larger groups (e.g. the Boston Red Sox), and the games are proxy for larger group competition and serve as ritual events that reinforce group identity.<sup>14</sup>

It is possible for individuals to differ in their grasp of the details of a frame and its significance. For instance, one sometimes experiences whispered discussions in the course of a synagogue service about how a particular bit of ritual is to be performed and why, whether doing it the way it is being done “really counts”, and what should be thought of someone who does that way. At the same time, many participants are largely oblivious to such hairsplitting distinctions.

It is also worth pointing out that our control of frames isn't airtight. There is often "leakage," as when competition within a game leaks out of the frame of cooperation and turns to downright hostility or even lasting enmity. The converse also occurs: bargaining and trading, although at bottom competitive, can lead to affiliative bonding.

This is of course all informal and descriptive. But I think it's a necessary prelude to asking the harder question, the one that properly belongs to cognitive neuroscience: what computational and/or neural mechanisms do we have to posit in order to produce this ability to frame and reframe recursively, and in order to permit the *learning* of this behavior? Team sports are grasped without effort by 10 or 11-year old children – though not by other primates, as far as we know. (Chapter 7 sketches a treatment of “action schemas,” a prelude to a more formal understanding of frames.)

#### 4.7. Rules and other normative principles

Let's think about a different aspect of games: the rules. Rules show up in many different domains. They take a general form something like this:

In frame F (or context C), you  $\left\{ \begin{array}{l} \text{should} \\ \text{must} \end{array} \right\} \left\{ \begin{array}{l} \text{do} \\ \text{not do} \end{array} \right\} X.$

Spelled out a little more explicitly we might get something like this:

---

<sup>14</sup>Shore 1996 (Chapter 3) offers an interesting discussion of many of these points, in particular the many changing layers of framing going on at once, in the context of baseball.

In frame F (or context C), if you {do/don't do} X, consequence Y of good/bad value to you will ensue.

Different sorts of rules differ in the kinds of consequences that they promise or threaten. Some examples:

- In games, the rules define a temporary space of action within which various rewards or penalties obtain. Breaking the rules incurs a penalty or breaks out of the frame.
- An obligation (or contract, including a promise) specifies certain actions that the holder of the obligation is to perform for the benefit of the person to whom the obligation is made. If I fail to meet my obligation to you, you get the right to perform some action that harms me. For instance, if I fail to pay off a debt to you, you have the right to demand restitution and perhaps further sanctions against me. Depending on the sort of obligation, you may be entitled to punish me yourself, or you may have to appeal to the group as a whole or to the group's designated authority to impose punishment on me. (Chapter 8 discusses obligations in much more detail.)
- A legal code designates certain actions as desired or sanctioned by the authority of the group (whether or not assented to by group members); the consequences of reward or punishment are carried out by designated representatives who act as proxy for the group.
- A system of moral or ethical rules designates certain courses of action as morally good and others as morally bad (and usually leaves yet others neutral). As far as I can see, the consequences associated with moral rules generally concern the approval and trust of community members. If you do something morally good, people think more of you and trust you more, and if you do something morally bad, the opposite. Trust in an individual, in turn, translates into increased interest in cooperative enterprises with this individual (Ellickson 1991).
- Religious codes replace approval by the community with approval by the deity or deities. In the Judeo-Christian tradition, the consequence isn't just approval or disapproval, it's specific reward or punishment, perhaps in the afterlife. Jewish tradition even sees its religious codes as a legal contract between God and the group.

One could go on and cite many other kinds of rules: parents' rules for their children, rules of etiquette, and I imagine you can think of others. I think, though, they are basically all of the same form; they differ only in the frames within which they're applied and in the general form of the consequences. A group's code of conduct is made up of such rules, explicit or tacit.

As noted above, the language used to express all these sorts of rules is pretty much the same, involving for instance the use of the modal verbs *should*, *must*, and *may* and adjectives such

as *right* and *wrong*.<sup>15</sup> As a consequence, they are not always clearly distinguished. For instance, moral/ethical codes are often conflated with religious codes. To be sure, aspects of religious codes often do state moral principles. But that does not make them the same. One can have a moral/ethical code independently of religion (think of honor among thieves, and perhaps desert traditions of hospitality); and many religious codes such as principles for performing rituals hardly fall in the moral/ethical domain.

Similarly, some of the literature on promises I've come across conflates the contractual and ethical domains. They treat the consequence of breaking a promise as disapproval by the community, that is, as an ethical breach. They fail to observe that there is simultaneously a contractual breach, which gives the individual to whom the promise is made very specific rights. Thus breaking a promise has a consequence in both domains. Similarly, legal contracts (as opposed to mere promises among individuals) have consequences in both the contractual and legal domains. (See chapter 8 for more discussion.)

A particular action may have conflicting consequences in different normative domains. A classic case is the evil landlord in the melodrama, who is foreclosing on the poor widow in exercise of his contractual right, but who in so doing is acting in violation of the moral code. In the other direction is nonviolent civil disobedience along the lines of Gandhi and Martin Luther King, which violates the legal code but is in conformance with what is taken to be a higher moral code. More generally, my sense is that at best, explicit legal and religious codes are intended as imperfect incarnations of a more inchoate sense of morality; at worst, of course, they can be used to legitimate the raw exercise of power.<sup>16</sup>

On the other hand, it may be important not to push the distinctions among these rule types too hard. For instance, Turiel's (1983) widely cited experiments show that young children readily distinguish those norms that we could "decide to change" – the social conventions – from those that we could not, i.e. genuine morality. A number of researchers following this lead (as well as the lead of standard moral philosophy, I think), have taken the view that morality should be studied in isolation and that, in order to get at the root of human nature, we ought to strip away

---

<sup>15</sup>Moreover, these are not the only uses of these words. There is a "prudential" sense, used for giving advice: *It's raining, so you should take an umbrella*. This sense can be distinguished from the "rule" sense by who benefits: *You should take an umbrella = It would be good **for you** to take an umbrella* vs. *You should write your aunt a thank-you note = It would be good **of you** to write your aunt a thank-you note*. The former is for *your* benefit, and is therefore prudential; the latter for your aunt's, and therefore normative. There is also a predictive sense of the modal verbs: *The bus should arrive soon; The bus must be there by now*.

<sup>16</sup>John Mikhail (2000) has stressed the way issues in legal codes mirror intuitive judgments of morality.

the relative superficiality of social convention.<sup>17</sup> I disagree with this stance for four reasons. First, I have the impression that cultures differ in what they themselves consider to be morality as opposed to social convention, particularly with respect to issues such as sexual mores and slavery. Second, when we ourselves look at a culture from the outside, what looks to us like social convention and what looks to us like morality are inextricably intertwined. Consider for instance the Ten Commandments, where alongside the moral dictate “Thou shalt not kill” is what looks like a social convention: “Keep the Sabbath”; nevertheless, according to the Book of Exodus, violation of either one is punishable by death.<sup>18</sup> Third, morality is deeply tied up with the logic of groups. A group may recognize its own particular variations on morality: “We hold ourselves to a higher standard,” or “If we do this, our gods curse us, but it’s different for you.” In addition, killing the enemy in a war is typically regarded as good, not as immoral at all (unless of course you are the enemy). The sociologist Jane Jacobs (1994) proposes in fact that human societies universally have two independent systems of morality, with partly contradictory tenets. One, the so-called guardian syndrome”, concerns how to keep the group cohesive and defend it against aggression from other groups. The other, the “commercial syndrome”, concerns how to participate in trade with other groups. The two moral systems coexist, if uneasily, in every successful society; the trick is to know when each is appropriate (an issue of framing).

But there is a fourth reason – a theoretical reason – not to isolate morality from social convention. One of the motives for such a move of isolation is to discover “genuine human nature” beneath the variations of culture. But such a view incorrectly takes the variations of culture to be superficial and uninteresting. Recall the analogy to the study of language. Universal Grammar is not a theory of what is universal in language, it is a theory of children’s ability to learn language. It has to take into account the range of variation among languages, what happens frequently and what *never* happens, what’s easy for children to learn and what’s difficult. Similarly, in seeking to discover the aspects of human nature underlying human society, we cannot just insist on the universals of culture: we should be looking at the range of variation and how the common issues of humanity play out crossculturally. Thus the system of norms as a whole seems a more ecologically appropriate object of study, particularly given the empirical considerations I’ve just raised.

One thing that intrigues me is the way normative rules -- of all sorts -- are taken to be objective entities in the world, albeit abstract. We face a certain cognitive dissonance. On one hand we know that people made them up. But on the other hand they're hardly imaginary! Within a game, I objectively win or lose. If I break a promise, or if I fail to pay my taxes, the consequences are real. So rules, once they are established as consensual, are practically as irresistible as laws of physical causality. In particular, rules that we call “moral” are conceptualized as timeless and universal, whether or not they really are from a crosscultural

---

<sup>17</sup>I take Mikhail to be taking this line, as well as possibly Stich and associates.

<sup>18</sup>Exodus XXI, 12 for murder, Exodus XXXI, 15 for working on the Sabbath.

perspective. This is why moral relativism is so repugnant to many people: they reason that if a rule is relative, it *can't* be moral. (Chapter 9, if I write it, will address this issue.)

Again, it would be of interest to explore children's understanding of rules. How do rules differ from pure principles of associated stimulus and response? Do children understand rules in the same way in all normative domains, how are rules learned, and so forth? Most of the rule systems I've mentioned purport some sort of impartiality or fairness. What do children of various ages think fairness is? Are there brain deficits that lead to failure to understand rules? And so forth. Building on early work of Piaget (1932), Kohlberg (1981/84), and Turiel (1983), an interesting body of work seems to be developing in this area.<sup>19</sup>

To be sure, the codes of conduct particular to a community have to be learned by children, as well as by outsiders who interact with or join the community. But it's quite possible that we don't have to learn *that there is such a thing as a code of conduct*. Rather, the pervasiveness of such organization suggests it is a skeletal conceptual structure around which humans organize their social existence.

#### **4.8. Distribution of resources: ownership, territory, transactions**

To be filled in: Every community faces the issue of distributing resources: food, fuel, artifacts, housing, land – and labor. What ownership is. What you can own varies from culture to culture (land? people?). Territory belonging to the group. Transactions: exchange based on value (+value of engaging in exchange). Principles of distribution: Fiske's 4 modes as context-dependent varieties of fairness.

#### **4.9. Where science bumps up against politics**

Many of the issues touched on here are very tricky -- and not just scientifically tricky. They run below the surface of a lot of intense public debate, not to mention thousands of years of philosophical and religious discourse. The underlying question is: What are the sources of principles of fairness and of moral/ethical values, particularly those that are conceptualized as universal and timeless? For instance, when in section 4.5 I spoke approvingly of efforts to teach tolerance, what justified that stance? Why shouldn't we instead applaud efforts to teach xenophobia and white male supremacy?

A great deal of western and especially American tradition has regarded moral values as given by God, for instance in Jefferson's phrase "endowed by their Creator with inalienable rights." I am given to understand that Islam takes a similar stance. Immediately upon the

---

<sup>19</sup>Harris, Kuhlmeier, Stich and associates, Mikhail, Dupoux/Jacob/Nurock, Blair, .... For a trenchant critique of the Kohlberg "stages", see Macnamara 1991. The Very Brief Version of his critique is that if young children had the view of morality attributed to them by Kohlberg, they would not be able to understand the point of fairy tales like Cinderella.

publication of Darwin's *Origin of Species*, the threat to this position from evolutionary theory was sensed by all participants in the debate, and certainly in the US it is connected with the rise of religious fundamentalism and its continuing hostility to evolutionary theory. For if morals are not given absolutely by God, where do they come from? If morals are relative or subjective, just made up by people, who says you can't make them up any way you want? How can you argue against nazism or communism -- or secular humanism or drugs or free love? Better to trust in what God says and use it to get what you want. This would be funny if it weren't so serious. The consequences for education, public discourse, and funding for science and especially the humanities are obvious to anyone. And even today we see military invasions being justified on grounds of timeless absolute God-given morality.

To my knowledge no one has offered a coherent answer to the question of how moral values are to be grounded within a society that does not rely on a particular God's authority -- that is, within the global society we all live in now. Perhaps utilitarianism or pragmatism a la Dewey is presented as an alternative, but its axioms too are left unjustified. Mahlmann 2003 discusses several modern positions that have concluded that the only possible foundation for law is the threat of force, and shows why they are unsatisfactory, and not just because of their unappetizing conclusion. For the most part I don't think that people opposing the religious fundamentalists and the economic Darwinists really try to answer the question; they just assert their own moral codes and point out the contradictions and vast helpings of self-interest in the right-wing position.

One folk theory, taken up by philosophers such as Locke, Rousseau, and Kant, has it that at some point people sat down and agreed on the "social contract"; this does explain the sense of social codes as joint commitments. John Rawls (1971) also takes up this idea, while making clear that he intends it as a fiction -- he proposes that we should think about a code of justice from the position "as if" we were hypothetically devising a social contract. Now formal legal systems are indeed developed by people sitting down and making them up, but I doubt this is the case with most elements of codes of conduct in most societies. Of course, the parallel (and robust) folk theory of *language* -- that people sat down and decided how to say things -- is totally implausible.

I am not so sure that evolutionary psychology can provide a proper grounding for values either, although perhaps it can offer some boundary conditions. Let me offer two contrasting examples of what I mean here. First: I think it's been well established by evolutionary psychologists such as Dawkins that there is an asymmetry between males and females in reproductive strategy: reproduction is a small investment for a male, a large one for a female. This asymmetry drives lots of behavioral asymmetries observed in lots of species. One particular game-theoretic consequence is that males are more likely (or more inclined) to be sexually promiscuous than females, a phenomenon we observe in humans as well. But we wouldn't want to argue from this biologically-driven logic that this is the way it *should* be -- that we *should* condone or even encourage male promiscuity. Morality ought to be properly distanced from biology here (notice that even here I have to say *ought*).

Now a contrasting case: One aspect of the logic of groups that I mentioned earlier is that social pressure from the group dampens aggression among group members -- the group protects its members from harm by other group members. Thus in our culture, the legal system punishes not only physical aggression like assault, but also economic aggression like stealing. Similar institutions are found in some form in every culture, even in the absence of written legal codes. In fact, Frans de Waal (1996) has observed that similar things happen in chimpanzee groups: dominant individuals often step in to break up fights.

Now it's a tenet of modern free market capitalism that businesses *should* be free to exploit people economically, and that the government *shouldn't* be allowed to protect citizens from such exploitation (that this "distorts the market"). Capitalism is of course based on bargaining as its basic form of interaction, which on my analysis a while back emerged as a kind of tamed aggression -- but aggression nonetheless. Globalization can then be seen as an attempt on the part of corporations to operate outside the group's constraints against aggression, as it were operating as lawless pirates. Thus we might conclude that the basic premises of free market capitalism flout the universal logic of groups, as established by the theory of social cognition. But if I take this conclusion as an argument against free market capitalism and globalization (as I did in *Patterns in the Mind*), then am I forced to accept sexism as justified too?

I don't think that a theory of social cognition can offer a full answer to the grounding of values; at best it can help us appreciate a fuller range of possible answers (this is the conclusion of Mahlmann 2003 as well). But I think it's important to remember that these political issues *are* part of the territory. If one works in this area, one ought to be prepared to discuss the questions openly and thoughtfully, bringing to bear our (hopefully) growing understanding of what sorts of cognitive entities moral codes are, of the role moral codes play in the functioning of a society, and of the innate underpinnings of social understanding that help shape moral codes in every culture. To be sure, our politics should not infect our science, and vice versa. Still, we never can be just innocent objective scholars. I think we have to be alert to potential political consequences of our research, and in particular we should be concerned that our work is not taken up by demagogues eager to make pernicious political points (as happened both with Darwinism and sociobiology).

All right. This chapter has been an extended meditation on big issues for a field of inquiry whose parameters are just beginning to fall into place. Many of the issues I've talked about, having to do with the constitution of society and the role of individuals in it, have been discussed by everyone from the Greeks through all the great religious thinkers and social and political philosophers. What I think is different in the approach here is that we have contemporary tools of cognitive neuroscience at our disposal, which can provide a far more comprehensive view of what it means to say there is a human nature. Maybe it's unfashionably romantic, but I do think that understanding human nature is the largest and most noble goal we can aspire to, and I am pleased to be part of the undertaking.

## References

- Bangerter, Adrian, and Herbert H. Clark. 2003. Navigation joint projects with dialogue. *Cognitive Science* 27, 195-225.
- Brown, Donald. 1991. *Human Universals*. New York: McGraw-Hill.
- Byrne, Richard W., and Andrew Whiten (eds.). 1988. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford: Clarendon Press.
- Cheney, Dorothy, and Robert Seyfarth. 1990. *How Monkeys See the World*. Chicago: University of Chicago Press.
- Cosmides, Leda, and John Tooby. 1992. Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, and J. Tooby (eds.), *The Adapted Mind*, 163-228. New York: Oxford University Press.
- Dawkins, Richard. 1989. *The Selfish Gene*, new edition. Oxford: Oxford University Press.
- de Waal, Frans B. M. 1996. *Good Natured: The Origins of Right and Wrong in Humans and other Animals*. Cambridge, MA: Harvard University Press.
- Degler, Carl N. 1991. *In Search of Human Nature: The Decline and Revival of Darwinism in American Social Thought*. New York: Oxford University Press.
- Diamond, Jared. 1997. *Guns, Germs, and Steel The Fates of Human Societies*. New York: Norton.
- Ehrenreich, Barbara, and Janet McIntosh. 1997. The new creationism: Biology under attack. *The Nation*, June 9, 1997.
- Eibl-Eibesfeldt, Irenäus. 1989. *Human Ethology*. New York: Aldine de Gruyter.
- Goffman, Erving. 1974. *Frame Analysis: An Essay on the Organization of Experience*. Cambridge, MA: Harvard University Press.
- Goodall, Jane van Lawick. 1971. *In the Shadow of Man*. New York: Dell.
- Grosz, Barbara J., and Candace L. Sidner. 1999. Plans for discourse. In Cohen, Morgan, and Pollack 1999, 417-444.

Hamilton, W. D. 1964. The genetical evolution of social behaviour (I and II). *Journal of Theoretical Biology* 7, 1-16; 17-52.

Herrnstein, R. J., and C. Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hobbs, Jerry. 1999. Artificial Intelligence and collective intentionality: Comments on Searle and on Grosz and Sidner. In Cohen, Morgan, and Pollack 1999, 445-459.

Mikhail, John. 20xx.

Pinker, Steven. 2002. *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Tooby, John, and Leda Cosmides. 1992. The psychological foundations of culture. In Jerome Barkow, Leda Cosmides, and John Tooby (eds.), *The Adapted Mind*, 19-136. New York: Oxford University Press.

Wilson, Edward O. 1975. *Sociobiology: The New Synthesis*. Cambridge, MA: Harvard University Press.

Zuriff, Gerald. 1998. Against metaphysical social constructionism in psychology. *Behavior and Philosophy* 26, 5-28.