Public Reason 1 (2): 45-72

Of Human Bonding:
An Essay on the Natural History of Agency

Mariam Thalos & Chrisoula Andreou[1]
University of Utah

Abstract. We seek to illuminate the prevalence of cooperation among biologically unrelated individuals via an analysis of agency that recognizes the possibility of bonding and challenges the common view that agency is invariably an individual-level affair. Via bonding, a single individual's behavior patterns or programs are altered so as to facilitate the formation, on at least some occasions, of a larger entity to whom is attributable the coordination of the component entities. Some of these larger entities will qualify as agents in their own right, even when the comprising entities also qualify as agents. In light of the many possibilities that humans actually enjoy for entering into numerous bonding schemes, and the extent to which they avail themselves of these possibilities, there is no basis for the assumption that cooperative behavior must ultimately emerge as either altruistic or self-interested; it can instead be the product of collective agency.

Key words: agency, altruism, bonding, collective action, cooperation, emotional attachment, identification, individualism, self-interest, team reasoning.

"Life is a long trip in a cheap car. In a dark country. Without a good map."[2]

The topic of what sustains cooperation among biologically unrelated individuals in evolutionary history knows no disciplinary boundaries. How do we explain the evolution and stability of such instances of cooperation? This question was heralded appreciatively in the present scientifically-minded era by E. O. Wilson in his controversial classic *Sociobiology*, although Thomas Hobbes was acquainted with a certain, unbiologized version of it much earlier. And the question remains with us still. Ranged among those who find it within the ambit of their concerns one can find psychologists, biologists, anthropologists, linguists, political theorists, students of computation, game theorists, and of course philosophers. The question has generated a wealth of cross-disciplinary conversation that promises to impact public policy palpably. For if it is determined that the correct answer to the question is that individuals cooperate (or comply with an unfavorable condition) only if doing so serves them severally according to a favorable pay-back schedule, then public policies will be framed accordingly: the success of public policies will be estimated according to the schedule of incentives they promise to complying individuals. This is already the foundation of many economists' proposals and theories – to the chagrin of those with more optimism about the better angels of human nature. But of course if the more pessimistic position on the question of how cooperation is sustained is in fact bet-

1] We are indebted to Lije Millgram and Nick White for thoughtful comments and conversations.

2] Opening words to Frederick Schick's *Making Choices* (1997), written with the aim of improving Bayesian decision theory.

ter supported by the evidence, the economists in question are not simply jaundiced but rather proceeding on the best possible grounds.

We shall argue that scientific theories contending that cooperation among strangers rests ultimately on a foundation of self-service, or service of one's lineage, are ill-founded, and indeed ill-supported by experimental evidence. For these theories are founded on an aprioristic restriction of the search space to mechanisms of launching behavior at an individual level. This restriction of the search space functions as a highly problematic assumption to the effect that motivation is an individual-level affair. We shall be challenging this dogma. To be sure, we are not the first to challenge the individualistic dogma, nor are we contending that all scientific theories on this topic are guilty of the individualism we shall be indicting. What we are offering here is a systematic criticism, across a range of literatures, and advancing also a corrective that helps to cast nonindividualistic proposals in a new light.

We shall be arguing that prior to the question of the evolution of cooperation is a parallel but different question that must be answered. This is the question of the evolution of agency itself: how did agency, as such, emerge on the evolutionary landscape, and in such units as we actually find it on the ground today?

Now, an agent is a unit – indeed a *unity* – that takes (or at least launches) action; an agent is related to its deeds as author and not merely or necessarily as proximate cause. To be sure, it is an empirical question whether a given thing, indeed anything at all, qualifies as an agent in this sense. For about a century now the wisdom among those with naturalistic inclinations has been that the idea of an agent is a construct that has no role to play in naturalistic explanations. Relatedly, much of psychological theory in the last century consistently treats the "self" as simply a collection of so-called *self-beliefs* or *self-attributions*, a collection of self-relevant beliefs, rather than as something that could legitimately be treated as an entity in its own right.[3]

There is of course a clear advantage to those conducting primarily psychological or behavioral science, as well as to philosophers following their progress, of conceptualizing agency in terms of performance criteria: this is that it is unproblematic to examine and document an organism's cognitive performance on tasks, enumerate and diversify the tasks examined, and then proclaim when once enough of these are co-present in an organism, that the organism qualifies as an agent (or at least as having a precise quantity of intelligence, that seems itself to be a place-holder for agency).[4] And it is quite possible that a critical mass of such capabilities will be sufficient to guarantee the presence of agency

3] See for example the essays arrayed in Wegner and Vallacher (1980) as well as the pieces in Duval, Silvia and Lalwani (2001), in which "self," as such, is never distinguished from "self-concept," "self-awareness" or "self-standards". Contrast this with new and important research on self-regulation prominently led by Charles Carver and Michael Scheier (2001), in which the "self" is construed as a sum total of self-regulation processes; and cf. contributions by Demetriou and Kazi (2001).

4] The strategy is illustrated admirably in Byrne (1995), but Byrne makes no contentions vis-a-vis agency as such.

as well. But it is very questionable to suppose that such a critical mass, despite being sufficient to guarantee the presence of agency in prototypes of the species, is necessary for possessing agency, or sufficient in every case. More problematic still is the supposition that a critical mass in ability to perform such tasks itself *amounts to* or *constitutes* possession of agency. The most obvious reason for denying this latter supposition is that execution of the tasks in question might be organized in a distributed or decentralized fashion, and that a disunified aggregation of performances, however expertly carried off, does not add up to an agent – which, by definition, is a unity.

This state of affairs is perhaps largely responsible for the sustained flowering of transcendentalism. Dissatisfaction with the naturalists' treatment of the agent has contributed to a growing sense that science does not – and cannot – give an adequate or complete treatment of agency, because the reality of agency, as such, *transcends* the methods of science. Transcendentalists defend the existence of a special sphere or realm of which science cannot treat: the first-personal, transcendental world of the Self and Others, which has at least since Kant been considered the private preserve of Philosophy with a capital P. Advocates of this transcendentalism usually insist that we *not* call the study of the transcendental sphere a science. These intellectuals do not suffer from science envy; they are card-carrying Philosophers. Science, according to transcendental philosophy, can take us only so far in the intellectual journey. And the suggestion is that perhaps Reason, with a capital R, or simply Intellect, can take us beyond the frontiers of science.

In our view, social scientific research on agency has tended to define its search space too restrictively, and there is no reason to deny that agency has a role to play in naturalistic explanations. Now of course there are important details to work out – most notably, handling the question of what qualifies an entity as an agent. And of course there is no shortage of answers to this question on the transcendental side. The trouble for our purposes is that a preponderance of published opinions on this subject, on both sides of the transcendental divide, simply assumes that the boundaries of agents coincide with their skin, or their fur or what-have-you.[5] On what foundation does this assumption repose? On nothing but philosophical (decidedly even political, and in particular neoliberal) dogma, as will become clear. As we will show, the answer to what agency consists in must be responsive to empirical findings about the behavior of contemporary humans, as well as to evolutionary considerations, and these empirical findings contradict the dogma of whose truth we are constantly being assured. The failure of this dogma, as we shall argue, is reason to be suspicious of current wisdom on the cooperation question. We will show that numerous incidents of cooperation among the unrelated are inexplicable by a calculus of self-service, and instead are better explained by a calculus whose subject is some "we." We will provide a taxonomy of "we"-s that distinguishes among its targets of analysis on the basis of how they are forged, which will provide a natural taxonomy for types of cooperation.

5] Indeed one of us has proposed an alternative conceptualization of agency that deliberately shuns this assumption: Thalos (2007; 1999), cf. also Thalos (2008).

Before we begin it will be well to handle in advance one obvious initial reaction to our proposal. Notice first that we couch our contention in terms of the entity level at which identification of motivation is appropriate. A critic might reply that motivation is in no way the issue here, that what is at stake is the level at which interests are served, or simply at which advantage accrues, and that any talk of motivation has been purely incidental or metaphorical, simply a rhetorical device for marking interests. Economists, evolutionary biologists, psychologists and ecologists – among others – have settled on what may at first glance seem a sterile or bloodless way of handling explanation of the individual human behaviors they aspire to explain. They view individual human behavior as best modeled on a fiction – namely, that the behavior is undertaken as a means to solving a certain decision problem, through maximizing a return on some investment. They do not view the individuals that manifest the behavior as *themselves* carrying out the calculation that solves the decision problem. Nor do they view these individuals as themselves understanding their situation in terms of a decision problem that (first) calls for a cost-benefit calculation, and (subsequently) leads to a motivation in favor of the option that wins the day. Therein lies the fiction: there is no real-time decision processing in the proper sense of the term; indeed, there may be no cogitation of any kind. Nonetheless they view the behavior as *best explained* on a model that weighs certain costs against certain benefits.

We grant the propriety of this form of explanation; in fact, it is one of our own contentions that the issue of agency is prior to and in many ways independent of psychological mechanisms that today underlie motivation. And in drawing attention to motivation we too are drawing attention to the level at which advantage accrues, contending that agency issues are intertwined with advantage and that evolutionary pressures can be brought to bear upon units of agency themselves.

Why then draw attention to the topic of motivation? Why not retreat to a *generalist* position on the subject of explaining behavior?[6] The generalist position seeks to explain behavior by directing attention purely to advantages conferred by the behavior on those entities engaging in it, without making any attempt to identify the relevant psychological machinery for controlling that behavior. The issue would then coincide exactly to the issue of levels of biological selection.

What the generalist proposal lacks in detail it makes up for as follows. The generalist proposal is simply to identify behavior that can outperform a range of competing behaviors (in evolutionary terms, during a certain period in evolutionary history), and subsequently to claim that this advantage makes the behavior inevitable even for those of us who come so much later. Advocates of this generalist position argue that a less generalist account, that deals in psychological details, might obscure this fact. It might suggest that details matter, when they don't. For if the behavior weren't achieved through the particular ways it was achieved, it would have been achieved some other way. And this is the generalist's point. We could subsequently cast our contentions in terms of the level at which

6] This is Robert Batterman's term (1998, 76-102) to refer to the work of Robert Axelrod (1984) and Brian Skyrms (1994, 305-320; 1996).

interests accrue, and leave out discussion of motivation entirely. That would serve certain of our immediate ends. But it is unsatisfactory in the long run to leave out issues of motivation, because in the end the agent is an entity defined at least in part by how it navigates in relation to motivations, its own and others'. And so to say that agency emerged in response to an evolutionary pressure is to say that unified and motivated entities so emerged. And with this statement comes the obligation to develop an account of how agency manifests itself in developmental time, and how it is transmitted down a lineage.

According to our account, motivation is a matter of being drawn to a goal or object – a feature abstracted enough from any physical realization of it to be realizable by a group or collection of dispersed biological individuals. A great deal hangs on what we count as agents (especially as increasingly much of what human life depends upon is determined by what happens to common-pool resources). And so any contentions about what may so qualify should be supported with argumentation and not merely subjected to fiat. This is the imperative to which this paper is responding. It will proceed in a way that treats agents as natural kinds, asking but not purporting any decisive answers as to how units of action can emerge in evolutionary time.

I. THE QUINTESSENTIAL SOCIAL SPECIES

"Sex," as E. O. Wilson (1975) remarks, "is an antisocial force in evolution." It constrains large-scale organization and division of labor – not because it interferes with labor or the individual variations among organisms that makes division of labor efficient, as quite the contrary is true – but because it stands in the way of unproblematic division (however inequitable) of the fruit of such labor. With the rare exceptions of monozygotic (identical) siblings, no two organisms in a species that reproduces sexually are genetically identical. Conflict – if only in strictly biological, reproductive terms – is therefore inevitable. To the extent that more for your interests means less for mine, collaboration between the two of us cannot be without its strains. Whereas, by contrast, where there is no genetic gap between the two of us, your interests and mine will coincide exactly, and there is no sense in which more for your interests means less for mine. Our cooperation in such an instance can proceed without hesitations.

Thus one way of overcoming obstacles to large-scale cooperation is to close the genetic gap, as has been done in the social insects of the order Hymenoptera via the device of *haplodiploidy*.[7] With sterile castes and suppression of reproduction among females, the gap

7] This is the mechanism by which males (developing from unfertilized eggs) have only one copy of each chromosome (haploid), while females are wrought the usual way and with two copies of each chromosome. This mechanism has important consequences: a queen's daughters from the same mating (called *supersisters*) are highly related to each other, and a female is more related to her sisters (on average 75%) than she is to her own daughters (on average 50%). Thus haplodiploidy opens the way for a worker caste, devoted to helping their mother. Sterility is a superior strategy when it is more expedient (less costly) to help a mother beget a sister (or many sisters) than to have a daughter of one's own. See Wilson and Hölldobler (1990) and (2009).

between individual interest and collective interest among the social insects is appreciably closed, making the family or colony the (veritably one and only) unit on which the forces of natural selection act. As R. A. Fisher remarked, "The insect society more resembles a single animal body than a human society…The reproduction of the whole organism is confined to specialized reproductive tissue, whilst the remainder of the body…tak[es] no part in reproduction" (1958, 200). In pronominal terms, social life among the social insects is a matter of "we" (at home) and "they" (when one colony encounters another). "I" and "Thou" have no genuine place.

By contrast the gap of genetic relatedness between organisms in a modern human community is as large as it can be among sexually reproducing species. Still, human societies enjoy (if anything) a run-away division-of-labor that leads to continuously escalating organizational structures entirely unprecedented on the planet. Human beings today live in large settlements, many of them phenomenally large, covering vast territories, and comprising numerous genetic lineages of unrelated individuals. According to Wilson's ground-breaking 1975 treatise, human societies comprise one of four pinnacles of social evolution, and cannot be explained entirely by mechanisms that support the welfare of close kin – mechanisms that ably explain the other three pinnacles of social evolution: colonial invertebrates, eusocial insects and nonhuman mammalian societies. As early as 100,000 years ago, humans lived in hunter-gatherer family units tied by cooperative bonds at a tribal scale, having no more in common than language and distant common ancestors. Even the simplest contemporary tribal societies link family units of a few tens to create societies of a few hundred to a few thousand, held together by common sentiments of membership, "expressed and reinforced by informal institutions of sharing, gift giving, ritual, and participation in dangerous collective exploits" (Richerson & Boyd 1999, 254). These tribal ties – very possibly constituting a necessary developmental stage along the way to large settlement living – are unprecedented in evolutionary history.[8] How is this form of social organization – which we will refer to as *network society* – to be explained?

Full-scale settlement living has many advantages: easy resource defense and reduced vulnerability to predation are perhaps the least controversial. So it might seem reasonable to propose that social living amongst humans, and the precursory network societies, evolved precisely because of or for these advantages. But there are problems with this proposal: for there also are disadvantages to large settlement living, with enormous susceptibility to disease being among the most important.[9] How do the advantages of settlement living measure up against the disadvantages?

8] Richerson and Boyd write: "We know of no close analog of tribes in other species" (2001, 211); they also write that "larger, more complex societies are generally able to dominate smaller, simpler tribal societies, and a ragged but persistent trajectory of social evolution toward ever more complex social systems continues to the present" (1999, 254).

9] Diamond (1999) powerfully documents the impact of contagious disease upon human societies in an important class of cases.

One suggestive strategy is to enumerate more and more advantages (and disadvantages too) to large settlement living – more than the obvious ones – and argue that together these constitute an enormously favorable balance over hunter-gatherer ways of life in nuclear family units. This strategy might work, if we could be assured that the evolution or development of the newly enumerated advantages does not require that large settlement living, and solutions to the problems and risks associated therewith, precede them. For example, it might be reasonable to suppose that settlement living makes possible the production of a food surplus, and so allows a large group to stockpile against the risk of famine. But does food surplus production require that large settlement living already be long and firmly established? If so, then we may have part of the story concerning the *expansion* of settlement living, but no part of the story concerning its *establishment*.

Eventually, the problem with the strategy of simply enumerating the advantages of large settlement living is that as we multiply purported advantages, ultimately we will run headlong into the problems of unrelatedness: what are the advantages that accrue to unrelated individuals through (for instance) the division of labor? For it might well be that the division of labor is more advantageous to some individuals than to others, for whom it might be a considerable burden. Unlike reduced susceptibility to predation, not everyone benefits to the same degree from the division of labor, since the surplus from it is rarely divided evenly. Indeed, some may not benefit at all – they might actually suffer in relation to how they might have fared with less social organization.

The point here is that the move to cooperative network living among the unrelated cannot be conceived entirely as a solution to a problem of pure coordination; for unlike solutions to problems of pure coordination, community living is not always win-win, or at least not obviously so. The mixed blessings of the likes of the division of labor, for example, are just as liable to destabilize the growth of settlements as they are liable to foster it. Thomas Hobbes knew something of this reality when he wrote that, in the state of nature, "there is no place for industry, because the fruit thereof is uncertain, and consequently, no culture of the earth" (1994 [1668], 76). To be sure, while the blessings wrought by a food surplus are blessings indeed, if I am not so positioned within the network society as to be confident I will be enjoying them, why should *I* count them as advantages of settlement living?

And so we arrive, finally, at a fundamental record-keeping question – the question to which we will devote the efforts of this essay: to whom is settlement living (or anything else) supposed to be advantageous? Must we view advantages as accruing, always and everywhere, to individuals, considered individually, or can our record books contain entries for groups, considered as groups over and above the individuals that comprise them? If the latter, then it might turn out that settlement living is a mixed blessing for both types of entries. As we will see, this is an important issue, as it serves to illuminate the question of just how an advantage functions in the logic of a purported explanation of its contemporary prevalence. We shall make room for answering the question vis-à-vis associations of

the unrelated in much the same terms as Wilson answers the question vis-à-vis associations of the very tightly related.

Whatever must be said about the balance of advantages to disadvantages, complex social organization in large settlements (as well as those things that come with it: domestication of plants and animals, large-scale food production, technological innovation of all varieties, writing systems and systems of communication generally) has grown up semi-independently more than once in the human lineage, and developed along a variety of different lines, supported by a variety of different social arrangements and cognitive structures.[10] So there must be a basis for it. And we'd like to understand what that basis is in as general terms as possible.

It is important to emphasize two facets to this theoretical problem. The first is largely historical and ecological: how did complex social organization arise among humans in the first instance? Unique features of species are typically the result of the niche the species occupies – its biogeographical location. What specific ecological *problems* did complex social organization solve for hominids? What are the ecological precursors and prompts? Did the organization in question arise in stages? If so, what were these stages, and what the nature of progression through them? Must all societies that reach this or a similar complexity in their social organization take this same route to it? (In other words, is this destination reachable only by one developmental route, or are there alternates?) This facet of the problem concerns the dynamics of the development of social organization, in relation to ecological conditions that favor or impede it.

The second facet of the theoretical problem is less historical and more ontological: it is concerned with the nature of the mechanisms and the entities that sustain and maintain the novel social organizations, and the boundaries that these things impose upon potential developments. What is it about the entities themselves, and the ways they function and develop, that sustains the structure in question, and constrains its further development and evolution? Are the mechanisms entirely biological? Or perhaps they are psychological in character? Maybe they are purely social, or a combination of all three. Are they present in their entirety in (some or all) of the individuals whose populations enjoy the organizational complexity? Or are they features of populations as populations?

Evolutionary game theorists are concerned with the broadly historical/dynamical questions. Theirs is the generalist approach. Other research, some of it conducted in the area of evolutionary psychology, is concerned with sustaining mechanisms.[11] Ultimately, the research in these two areas must be synthesized. For if, on the one hand, we go without a treatment of the ontological problem, it would appear that we have no more than

10] Nisbett (2003) argues that East Asian cultures differ from western cultures in profound ways that are anchored in cognitive strategies.

11] Barkow, Cosmides, and Tooby (1995); Buss (1994). Other sources are: Buss and Malamuth (1996), and Crawford and Krebs (1998).

mythologies or just-so storytelling.[12] If, on the other hand, we go without treatment of the historical/ecological problem, we do not have clear confidence that the proposed mechanisms explain what needs explaining, and whether they are indeed the best candidates for explaining it.

Though it might be too ambitious to address the two facets of the problem simultaneously, each facet can be addressed with some sensitivity to the other. Our purpose in this essay is to enlarge the toolkit for modeling the evolution of cooperation on the ontological side of the question, by taking steps to develop of a taxonomy of bonding schemes among organisms. Via bonding, a single individual's behavior patterns or programs are altered so as to facilitate the formation, on at least some occasions, of a larger entity to whom is attributable the coordination of the component entities. Some of these larger entities will qualify as agents in their own right, even when the comprising entities also qualify as agents.

II. THE MULTIGENERATIONAL DIMENSION: IN THE BEGINNING

Why do organisms bother investing in reproduction? The answer, of course, is that those who fail to do so die without descendants, and only those who manage to do so maintain a lineage. Among plants, those that manage to invest resources into reproductive functions win; and similarly in the animal kingdom. But among members of the animal kingdom, some young require more than just the opportunity to live. They cannot simply 'take it from there'; instead, they require some care to actualize their reproductive potential. Parents who fail to provide such young with the necessary care, leave no descendants into third and fourth generations. So how could species with such young have evolved in the first place? Nature, as we will now discuss, has provided these organisms numerous ways of bonding with their young.

Mammals give birth to live young. And so mammal mothers can meet their offspring face to face, without much risk of mistaking the relationship. This provides an opportunity that nature can take advantage of to reward those mammals that provide an optimum of postpartum care for their young. Mammal young need to be nursed for a period of time; mammal mothers who fail to nurse their young run the risk of leaving no progeny behind. Those who provide normally do so by way of having bonded with those young. The more bonded, the more they provide. Homo sapiens, like all other mammals, some birds and some reptiles, are a bonding species. Bonding is a solution to an evolutionary

12] One profound philosophical worry about purely dynamical approaches to the ecological problem is that they might be largely irrelevant. The concern is that the origin of any phenomenon consists in a trajectory of unique historical events, and so it is subject to empirical inquiries – anthropological, archaeological and sociological. An investigation of purely evolutionary-dynamical issues may reveal that certain historical sequences are ruled out, but it's not likely to reveal a unique sequence as the actual one. And so anthropological, archaeological and other empirical inquiries would still be required to discriminate the most likely among eligible historical trajectories.

problem – indeed the most fundamental one. Not surprisingly then, bonding is a universal of mammalian life, and must come on the scene long before you and I are considering whether to live together as cooperating strangers.

This truth has yet to be integrated into explanations concerning how cooperative living amongst unrelated individuals comes to prevail, and how it is sustained and multiplied. Yet the answers to the questions of (1) why (and how) organisms provision their young, and (2) why (and how) unrelated or distantly related organisms provision one another, are fundamentally related. They are related as species of the same genus, as our proposal will explain.

It is now commonplace to view the structure of the ecological problem faced by would-be cooperators as a Prisoners' Dilemma (PD). Early explicit expressions of this idea are due to John Maynard Smith (1982), Robert Axelrod (1984), and William Hamilton (Axelrod and Hamilton 1988), who introduced the apparatus of game theory into biology, and evolutionary biology in particular. But germs of the idea go back as far at as Thomas Hobbes. This game-theoretical conceptualization is combined with the assumption that cooperative living amongst unrelated individuals cannot be explained in essentially the same way that parental investment in mammals is explained – namely, via an appeal to bonding.[13] This assumption is, we will argue, mistaken – bonding is a key to cooperative living even amongst unrelated individuals. This suggestion is bound to seem implausible if one supposes that all bonding takes the form of emotional attachment; but emotional attachment is just one of a variety of species of bonding, as we will now explain.

III. BONDING IS A GENUS

In this section we shall be offering a taxonomy of bonding. Before we offer our three primary taxa, we will do well to mention some forms of association that might, under the right circumstances, qualify as precursors, if not also as bonding taxa in their own right.

Coral polyps live in colonies. Each polyp benefits, perhaps in only a very small way, from the proximity of the others. Perhaps the sheer number of cohabiting polyps helps to create a more stable habitat, moderating to some extent environmental variables like temperature. Assuming this is all that the colony provides the individual, and assuming that coral polyps have to affix themselves somewhere and together is just modestly better than apart, this seems a rather flimsy basis for thinking there is something that deserves calling coordination of behavior here. Still, because polyps possess little or no locomotion, it would seem that their "choice" of a home is everything in the world that they are entitled to calling behavior, and so it might qualify after all. If in the end we view this example and

13] Recently Skyrms (2003) has added the stag hunt game as a conceptualization of the ecological problem. His model is also founded on the assumption that cooperative living amongst unrelated individuals cannot be explained via an appeal to bonding.

others like it as continuous with the others, then we shall perhaps wish to add another taxon to our taxonomy. But we shall leave that issue open for now.

Symbiotic relations are everywhere. In mutualism, a dyad is formed whose members are generally of different species, each benefit from a close ecological association. The association might be negotiated by a coordination of behaviors, or simply as a result of happenstance that each finds itself in the right place at the right time. If it is through the former route – through coordination of behavior in one of the three ways we will be discussing – it will make sense to refer to the dyad as a bonded entity.

Consider now the following three species of bonding:

Imprinting

Two individuals can become a type of natural unit via the process of imprinting. Goslings, for example, imprint on the first moving creature they see (usually their mother), and then constantly follow this creature around. To take a somewhat different example, when a mother ewe gives birth, she imprints on the smell of her baby's wool while licking off the amniotic fluid covering the newborn. Within five minutes "the door to maternal tolerance slams shut" and the mother rejects any baby that does not smell exactly like the baby (or babies) she has imprinted on (Hrdy 1999, 158).

Given the right conditions, imprinting effectively attaches mother and offspring, which helps explain why it is sometimes selected for. The point is simply that the behavior pattern is triggered by a cue and is from that point forward comparatively rigid. This fact explains the imprinting "errors" that have been documented: The first moving creature a gosling encounters may not be its own mother, but a curious researcher, and bonding might occur anyway. And a mother ewe can be 'tricked' into adopting a lamb that is not her own if the lamb is smeared with her fresh amniotic fluid.

Emotional Attachment

A different, more familiar form of bonding for us humans is bonding via emotional attachment. Bonding via emotional attachment is more gradual and less mechanistic than imprinting. It typically requires, and is reinforced by, extended contact or 'face time.' In evolutionary terms, bonding via emotional attachment has its advantages. Like imprinting, emotional attachment can serve to bond kin to one another. But unlike imprinting, it admits of degree, and thereby allows for commitment to vary with emotional investment. (This can result in a variation of attachment strengths, depending upon expected payoffs of the attachment.) For example, while a mother ape will not care for a newborn that cannot cling to her, she will care for an infant to which she has become attached even if it becomes too weak to cling. (This makes sense, as the evolutionary value of a sick 6-month-old is considerably higher than that of a newborn, provided its chances of recovery are good.)

But again there is room (and in fact more latitude here) for copying errors, because even in the absence of blood kinship and potential for reciprocity, contact can breed emotional attachment. And so sacrifices for genetically unrelated individuals can be rampant in environments in which contact between genetically unrelated individuals is rampant. (In one clear instance of this in our own society, emotional attachment to adopted infants explains investment and even sacrifice between the genetically unrelated.)

Identifying

Another familiar form of bonding for human beings is bonding via identification. Our cognitive capacities allow us to recognize bonding and to think in terms of "we"; relatedly, they allow us to theorize or postulate "we"-s and to act accordingly. In some cases, the recognition of similarities suffices to get individuals postulating a "we" and thinking in terms of their (collective) good (Kramer and Brewer 1986); this, in turn, may be enough to sustain their existence as a "we." When it comes to identification, what is logically essential is not the development of certain emotions or *feeling*, but rather the development of a certain *conception* of things.

Note that identification serves to unify not only groups, but also individuals. (Indeed, development of the first personal "I" is perhaps the first appearance of bonding of this kind.) At least typically, when one takes into account how one's current actions will affect one's future self, one does this out of one's identification with one's future self. Although one sometimes reasons from the point of view of oneself-now, one usually reasons from the point of view of oneself-as-a-temporally-extended-being or from the point of view of some "we" of which one is part. Such reasoning involves identification, but need not involve emotion; in particular, it need not involve a feeling of sympathy. For example, it is my identification with my future self, rather than a feeling of sympathy for my future self, that gets me to the grocery store on Saturday mornings regardless of whether I am feeling hungry.[14]

Notice that it is possible to postulate a "we" and identify with others based on a faulty conception of one's relationship with those others. One may, for example, think of oneself as part of a team, even though no such team exists because the other supposed team members do not think of themselves as part of a team. But even when one correctly thinks of oneself as part of a team, this may seem like a mistake in evolutionary terms (some will insist) if those with whom one identifies are, by and large, not kin. Such behavioral developments should not (or so one might argue) be carried forward into future generations. For identification supports sacrifice, and sacrifice for non-kin is not generally fitness-enhanc-

14] Nagel (1970) argues that reason requires that agents take the interests of their future selves and of others into account. Given the notion of identification, one might interpret this as suggesting that agents must (on pain of irrationality) identify with their future selves and with others. This position is bolder and more questionable than our own. All we are suggesting is that agents (at least human agents) *can* identify with their future selves and with others.

ing. And even if identification can be fitness-enhancing in our own era, where advantages can accumulate to identifiers in numerous ways, how can it have evolved in the first place?

IV. WHY ARE WE HYPER-SOCIAL?

In this section we shall draw together the main elements of our proposal, contrasting it with our competitors' while drawing the relevant parallels between bonded groups and individual organisms. The question we are trying to answer is of course: Why – and how – are we humans so hyper-social? The major competitor to our proposal is this one: We are hyper-social through reading each other's minds (this answers the "how" question; Baron-cohen 1995, Tomasello 1999), and this is so that (now turning to the "why" question) each of us can keep from falling behind in strategic advantage in a competitive world characterized by shifting alliances (Byrne and Whiten 1985). And so it's all for individual advantage. Less starkly put, because each of us has the capacity for understanding the minds of others, through being able to put oneself cognitively and emotionally/motivationally in another's place, we have been able to learn from one another, play with one another, share information and ultimately live together in (apparently) cooperative communities.

This proposal rests on an apparent consensus, forged among philosophers and scientists of cognition, around the Representational Mind, which is the hero of the so-called "Machiavellian intelligence hypothesis" (Byrne and Whiten 1985). The ecological challenge, on this proposal, is to predict the behavior of others, and this requires reading their minds so as to represent it to oneself in a kind of practical syllogism performed third-personally. Once one's mind becomes readable, one has to raise the stakes by learning to conceal one's mind as well as to penetrate through such concealments as others find advantageous to perpetrate, in an escalating arms race of deceptions and unmaskings.

One version of this proposal (due to Hrdy 1999, 2009) declines to some extent the "Machiavellian intelligence" aspect of this hypothesis: No, it is not for warfare's sake (either for concealing one's intentions or for unmasking others' intentions) that we understand each other's minds; it is rather so that each individual among us (adult and immature alike) can find and/or please *allomothers* – adult care providers and protectors, which are indispensable because maternal commitment among humans is much more conditional than among other primates; but, once again, it's all for the sake of individual profit.

One challenge to the entire project of Machiavellian intelligence seeks to provide an alternative model of cognition to explain all the phenomena of hypersociality, without reference to hyperactive Representational Minds.[15] For instance, Strum, Forster and

15] Behaviorists and cognitivists alike, pervasively both within the discipline of psychology and outside it, have used the term "agent" to refer to any entity with the representational properties of mind, without giving due consideration to whether there is an open question as to the connection between goal-orientation on the one hand, and representational states on the other. This is evident in the title of an essay – indeed, an entire volume of essays – by the prominent philosopher of biology Kim Sterelny, *The Evolution*

Hutchins (1997) argue for a distributed cognition model of social processing, seeking to "undermine the very narrow individualistic language of tactics and strategies and the disembodied view of cognition that have been the basis of our approach to the primate mind in the new cognitivist era" (1997, 73). But even here there is no forceful challenge of the notion that the gold standard for explaining behavior in an individual is *individual* advantage. To be sure, there is a suggestion that a larger "cognitive system" may be in evidence. But there is no suggestion that natural selection acts on characteristics of *it,* favoring some characteristics and not others.

Our proposal, by contrast, is this: We are hyper-social through having managed to bond in multiple ways (answering the "how" question), which has led to the adoption of multiple collective goals which result in an overall reduction in conflict of interest; and all this is so that (turning now to the "why" question) the units that have bonded together can outperform other competing units. This, in broad outline, is the analysis we find E. O. Wilson and Bert Hölldobler (both 1990 and 2009) making vis-à-vis cooperation in the order *hymenoptera.* And correspondingly, we make no commitments as to the physical realization of goal orientation among the bonded – whether via a Representational Mind or something else – just as Wilson and Hölldobler makes no such commitment as to the physical realization of superorganismic bonding among ants, although they are keenly aware of the existence of chemical signaling among them (and motor programs among honey bees). We believe that the issue of what is in an individual's mind in a moment of hyper-sociality, just as the issue of the chemical signals or dances exchanged in ant and bee communication, is in fact a red herring. What is of significance is the *communication-structuring* motivation. And this may be nowise evident in human minds as such, just as the analogous articles may be nowise evident in the chemical signals exchanged by ants.[16]

Wilson and Hölldobler (2009, 6-7) describe the construction, in natural history, of organism and "superorganism" as perfectly in parallel. Just as there is no question as to what is in the "mind" of a single cell when it contributes to the behavior of the organism it helps to compose, so also: "Nothing in the brain of a worker ant represents a blueprint of the social order…Instead, colony life is the product of self-organization…The assembly instructions the organisms follow are the developmental algorithms" (2009, 7). And these are simply the result of natural selection operating in the usual way, but on the superorganism as a whole.

---

of Agency (2003), in which the only entities that the title could possibly be naming are organisms purportedly in possession of folk-psychological states of believing and desiring; nowhere in the book does Sterelny acknowledge a need to give further account of agency, but takes it simply for granted that "agency" and "thought-out behavior" are co-referential.

16] In accordance with the observations in the previous footnote, philosophers who cannot separate the notion of agency from that of a compendium of representational states – as cognitivists nowadays cannot – will be inclined to respond here that, accordingly, we must not be going on about anything recognizable as agency. We reply that, to the extent that the concept of *motivation* is itself inseparable from that of agency, we are indeed going on about something recognizable as agency. And we remain neutral here as to the relationship of motivation and mental states.

Like Wilson and Hölldobler, we view bonding as the result of natural selection operating in the usual way, upon characteristics of the higher-level bonded entities themselves, rather than upon the characteristics of the composing individuals therein bonded. If this bonding proposal is correct, it might explain a great deal – some of which cannot be explained at all by competitor proposals. In the following section, we shall, by drawing on empirical work, elaborate on those phenomena which our bonding proposal is better-placed to explain; and in subsequent sections we shall further contrast our proposal with other bonding-like proposals.

V. IDENTIFICATORY BONDING BREEDS COOPERATION IN THE LABORATORY

Identificatory bonders are capable of assimilating themselves into larger "we"-units. And, insofar as they are capable of reasoning and acting, they are capable of reasoning and acting as part of larger "we"-units. They can be found asking themselves "What should we do?" and "What would be best for us?" (These are all too familiar questions within the context of family units.) When such questions arise, it is often obvious that *we* will do best if each of us cooperates with the others, and so our deliberation is focused *not* on *whether* to cooperate with one another, but on which of the options available to us as cooperators to aim at. And importantly: when an individual acts as part of a "we", that individual is not acting simply on reasons had as an individual – that individual is acting out of reasons had as a part of a larger entity.

Bonds can, of course, vary in strength (as they do for example in molecules). Where bonds are very strong, bonders will reliably make significant sacrifices in order to do what is best for the group. Where bonds are weak, bonders may be willing to make only small sacrifices for the group's sake.

Consider the following set of experimental results described in Dawes *et al.* (1997). In one experiment (run by A. J. C. van de Kragt, R. M. Dawes, and J. M. Orbell with S. R. Braver and L. A. Wilson), groups of subjects were put in PDs. Some of the groups were given ten minutes of discussion time before each participant had to decide what to do. Other groups were not given the opportunity to communicate. Each participant then gave her confidential decision to the experimenters before leaving. While the average cooperation rate in the groups with discussion time was about 80%, the average cooperation rate in groups with no discussion time was about 40%. These results led to the hypothesis that group solidarity might be very important for reliable cooperativeness. Two follow-up experiments sought to determine whether conscience or the opportunity to make commitments, rather than solidarity, could account for the increased cooperativeness among groups with discussion time.

In one of the follow-up experiments (run by Orbell, van de Kragt, and Dawes), all participants had some discussion time; but in some cases, the discussion was with the group that would determine the participant's payoff and benefit from any contribution

she made; while in other cases, the discussion was with participants whose decisions would not affect her payoff or benefit from any contributions she made. The idea was that

> if discussion triggers conscience, and our contributing subjects are acting to satisfy its demands, then discussion should enhance contribution to strangers. If, however, discussion elicits caring about group members, then it should enhance contributions only to people in the group with whom one interacts. (384-5)

The researchers found that, "contrary to the clear conscience hypothesis," discussion "does *not* enhance contribution when beneficiaries are strangers" (387).

In another follow-up experiment (also run by Orbell, van de Kragt, and Dawes), groups were monitored and any verbal commitments made were tracked. Researchers found that, except where every member in the group made a verbal commitment, there was no relationship between promising and actual choice. It was concluded that, while promises are effective in "universal promising groups," it is "solidarity – not commitments per se – that leads to the higher level of cooperation in [these] groups" (389).

Note that, as Dawes, van de Kragt, and Orbell point out, although group interaction can effectively generate group solidarity, interaction between group members is not essential for the generation of group solidarity. It has, for example, been demonstrated that individuals tend to exhibit solidarity with respect to individuals whom they do not know (and whom they have not had any contact with) but whom they think of as sharing their fate (Kramer and Brewer 1986).

The suggestion that rational players in a true PD might rationally cooperate is, to certain prominent game-theoretic minds, little short of apostasy. It can only proceed, as Ken Binmore writes, from "the wrong analysis of the wrong game" (1994, 114). For, according to Binmore, it follows from the very meaning of the notion of 'payoff' taken together with the fundamental game-theoretic conception of rationality that a rational player in a PD must choose to defect (and hence if a rational player does not do so, it was not a PD in the first place). The idea that players in a PD might give some pride of place to *sums* of payoffs, adding in payoffs not reflective of their individual concerns – and suggestive instead of some "we" – would suggest to Binmore and others that players in the dilemmas construed by Dawes and Orbell do not perceive themselves to be playing a PD. (To a first approximation, they are construing the situation as an iterated PD.)

To explain why players in real-life experiments (both in and out of the laboratory) frequently choose to contribute to a collective good when faced with dilemmas whose material payoffs accord with the PD,[17] many strategies have been devised that involve the transformation of material payoffs into utilities whose structure deviate from the PD sufficiently as to make the observed behavior consistent with the game-theoretic solutions.[18]

17] A meta-analysis in Sally (1995, 62) reveals that summing across all 130 PD experiments carried out between 1958 and 1992, the proportion of subjects choosing cooperation over noncooperation is 47.4 percent.

18] A useful cross section of this literature is surveyed in Hargreaves-Heap & Varoufakis (2004, ch. 7).

Specific implementations of this include ascribing to players a dislike of or aversion to inequality, or ascribing to players a liking for reciprocation (or retaliation); either way, the result is a boost in subjective attractiveness, to the relevant players, of cooperative outcomes.[19]

What critics of cooperative solutions to the PD as such are unwilling to acknowledge is this reality: when we put subjects in a PD-like situation, by fixing their payoff schedule accordingly, we aren't specifying that they must construe their dilemmas as dilemmas for individuals or dilemmas for groups (nondeliberating, to be sure, and possibly also scattered in space and time). Indeed we cannot do so: it is an experimental manipulation that is simply out of reach. But if participants elect the latter construal, neither decision theory as such, nor any of its axioms individually, can censure them for it. So, if subjects are indeed irrational to construe themselves as groups in such circumstances as we place them, it is not decision theory that can convict them of it. Is such a construal irrational? Perhaps, at least in certain cases, natural selection might be in a position to impeach them for a misstep in construal – or at any rate to punish them. But its verdicts are not to the effect of "irrational!" For "Team Think" is not – as such – irrational, even if it might be in some other way practically inadvisable under certain circumstances.

Michael Bacharach has been developing "Team Think" analysis of cooperation in PD cases.[20] On his analysis, a player in a PD can either construe the dilemma as an individualist or else favor a collective construal of it. But on Bacharach's analysis, this is not a further choice – so that the collective computation is no intermediate step towards an ultimately individualistic construal. On Bacharach's account, "team reasoning" is fundamentally opposed to individual reasoning, so that if you are conducting deliberations in the one idiom, you cannot rationally be conducting it also in the other. Any given concrete dilemma is "spontaneously" framed *either* as an individual choice or as a collective one – never both at the same time. But frames can vary. Someone, on his view, reasons as a team member if she chooses the act (if this is unique) that is her component of the profile that (as she has worked out) is best for the objectives of some group (1999, 32). This reasoning, as Bacharach maintains, "is a basic decision-making proclivity of mankind; … it is fundamental to the workings of organizations of diverse forms; … it is a concomitant of group identification;…and … it completes the theory that group identification is the basis proximal mechanism for successful human group activity" (2006, 121).

19] For specifics, see for example, Fehr and Schmidt (1999), Bolton & Ockenfels (2000), Geanakoplos et al. (1989), and Rabin (1993).

20] Unfortunately, Bacharach is not clear on the point of whether team reasoners in situations where their payoffs can be characterized by a PD matrix are really in true PDs. His untimely death in 2002 was a tragic loss to the discipline, as he was still in the process of writing a book that pulled together the themes on foundations of decision theory that he had been developing for decades. Fortunately he had completed enough of this book that we can get a picture with broad strokes (but which is ambiguous on the point Binmore stresses). Thanks to Robert Sugden and Natalie Gold for taking on this important work: Bacharach (2006).

For our purposes, it doesn't matter whether team analysis of the PD schedule of payoffs is conceptualized as fundamentally inconsistent with individualistic analysis. What matters for our purposes is whether the process of team analysis brings into existence a new agency that vies for recognition, both in nature (for the attentions of natural selection) and by the apparatus of decision theory. We think that the birth of such an entity is a fundamentally important evolutionary innovation. It fills important gaps in the evolutionary history of cooperation.

Like Dawes, *et al.,* and in company with Bacharach, we have cast our construal of ecological challenges in terms of units of agency ("I" and "we"). Another way of framing the debate might be in terms of altruism. There has been considerable discussion of the cooperation question as a question of altruism versus self-interest, with altruism being cast as the antithesis to self-interest. We believe that the selfish/altruistic dichotomy is misguided because it elides the difference, on the nonselfish side, between acting as a member of a collective and acting in the interests of another entity, as an entity separate from it. In other words, the selfish/altruistic dichotomy neglects as distinct another and possibly quite potent and pervasive form of motivational organization: the individual acting as a part of a collectivity. When someone votes against their individual interest, either as a private citizen or as a member of a governing body, is that person performing an altruistic act? And when a person bothers to perform what they view as an admittedly minor civic duty, for example by exercising a legal right to vote, is that an act of altruism? We believe there is an important difference between *civic*-mindedness and bona fide *other*-mindedness. And this is the difference that our taxonomy of agentic forms attempts to capture. In our view, the most probing distinctions are made in terms of agency: X acts on behalf of Y, where X ranges over all entities that can serve as agents and Y ranges over all entities to whom an interest can be attributed.

When the capacity for bonding is impaired, the consequences are liable to be pervasive; both one's sense of community and one's sense of self suffer, because each (according to our proposal) is achieved through a very general capacity for bonding. With this conceptualization in place, it is easy to see how someone can be both notoriously intelligent and manipulative, and yet also imprudently impulsive. This description fits perfectly the classic characterization of the psychopath,[21] though the dichotomy between altruism and self-interest has made it tempting to downplay the psychopath's imprudent impulsiveness and highlight his indifference to others instead. With our construal of bonding as an elemental human capacity for both transacting and conceptualizing social reality, it becomes clear that the elements of the classic characterization of the psychopath are not in tension. Quite to the contrary, it is natural to find them going hand in hand. Insofar as

21] The modern conception of psychopathy was articulated by Hervey Cleckley in his classic work *The Mask of Sanity* (1941). According to Cleckley's criteria, a psychopath is intelligent, manipulative, irresponsible, impulsive, inadequately motivated and entirely devoid of shame.

psychopaths suffer a general deficit in the capacity for bonding, they are as prone to harm their future selves as they are to harm other living things.

VI. DIVISION OF LABOR WITHOUT HAPLODIPLOIDY: OFEK'S BAPTISM BY FIRE

Haim Ofek (2004) offers important new insights, and an ingenious argument for the evolution of division of labor without haplodiploidy. It has been customary for some time now to consider the prospects for cooperation furnished by the promise of so-called *public goods* – goods like protection from predators, which are such that if provided to some are effectively provided to all in the group (nonexclusion property), and whose enjoyment by one more does not detract from or lessen its enjoyment by the originals (nonrivalry property). Hence game-theoretical research has modeled evolutionary challenges as PDs, wherein individual interest enters into conflict with that of the group because no one individual has an individual motivation to contribute to the creation of the good if no one can be excluded from its enjoyment once it has been produced. This obstacle to cooperation is commonly known as free-ridership, which (as common wisdom has it) is responsible for underproduction of the good in question.

Ofek, by contrast, draws attention to the creation of *contrived goods* as providing the incentive for division of labor. He brilliantly discusses big game and the domestication of fire as instances of contrived goods. Contrived goods (like public goods) are nonrival, in that the enjoyment by one more does not detract from its enjoyment by the originals, but it differs importantly in the dimension of exclusion: a contrived good is something that can be withheld from a noncontributor. Thus there will be a strong incentive to "specialize" in production of those things that cost you next to nothing to produce for one more consumer on the margin, and which is of some considerable value to that next consumer provided he doesn't already have some of it, if he has something of value (to you) to exchange. In circumstances where exchange is possible and not (in itself) costly, division of labor will be more efficient. (We will discuss at further length below what efficiency amounts to.)

How, specifically, can fire take the form of a contrived good? First, it offers prospects of exclusion when its production involves investment or skill – actually a suite of three separate skills is required that can be mixed for efficiency depending upon availability of fuel and other resources – specifically: incendiary, containment and maintenance skills; but the limiting resource is ignition (chs. 9-10). As Ofek writes, "The enormity of this requirement is no longer fully appreciated by modern humans within easy reach of matches. But until a point not so distant in the past it still posed a major challenge to all fire users giving ample opportunity for exclusion" (151). Second, fire's capacity for self-generation offers its human handler the prospect of being able to make fire with fire at no extra cost.

Absent the ability to make fire on demand (which is arguably a much later development in evolution), humans had to make opportunistic use of what can be borrowed from

nature (from natural fires started by lightning, for example). And here now is Ofek's ingenious argument: once "borrowed", there are a number of scenarios for how to extend the use of this recruited resource. There are three likely scenarios. First is what Ofek calls the "campfire" scenario – the communal fire. Fires have an optimal size: too small and they die out easily, too big and they are fuel-inefficient and hard to contain. This poses a large question of social organization: a central fire open to all the elements has all the features of a public good, and so subject to free-ridership:

> However beneficial to society as a group, individuals willing to undertake the painstaking task of tending a continuously burning central fire, providing it with fuel, protecting it from the elements (and from human error) – strictly as a voluntary act – are not easy to come by… To rely on voluntary action for the purpose of the day-to-day provision of a routine service in the mundane arena of subsistence, is to expect slightly too much of the wrong species in a wrong setting. (159)

Second is the "private fire" scenario: continuously burning small fires maintained individually by "private" users, typically family groups. The problem with such a system – and this is now the key to Ofek's argument – is accommodation of the occasionally unlucky or undervigilant user whose fire dies out. Short of waiting for another wildfire, the unfortunate's solution would be to "borrow" from a neighbor who is more lucky or more vigilant. Donation to the unfortunate might at least initially seem a costless gesture to a desperate mendicant. And common knowledge that provision to the unfortunate is available might seem like cheap 'fire insurance" to all. But, or at any rate so Ofek argues, this situation is inherently unstable: why be vigilant in maintaining your own if borrowing fire is free and maintaining is expensive? This arrangement too, it seems, is subject to the miseries of free-ridership, so it cannot maintain itself:

> if everyone can borrow fire on demand, it is no longer in anybody's interest to undertake the painstaking task of maintaining an ongoing flame. If the players could not figure out their own interest for themselves in the short run (say, because they presumably lack rational behavior), natural selection would figure it out for them in the long run. Either way, one ends up with a system comprised exclusively of borrowers, no donors, and no fire. (160-1)

Ofek's preferred scenario is the "incendiary hub", which calls for a firekeeper (someone who specializes in fire maintenance and control, but who is rewarded for this service by those who can borrow from him or her). This is a much more efficient system: cutting the cost of fuel and labor, and if there is more than one hub, the firekeepers can provide insurance to each other either freely or at the price of a small favor to be returned. And the division of labor is off and running! An ingenious argument – but flawed. The defect lies with the key move vis-à-vis the instability of the "private fire" scenario. Effectively, this scenario is *not* doomed by free-ridership. Here is why. If I am a vigilant firekeeper, and others are not so vigilant, I will soon find myself more often donating than receiving donation. And I might be noticing that the beneficiaries of my generosity are faring better than I am. Does this incentivize me to ease up on my diligent efforts? By no means! Surely I

will instead find reason to redouble my efforts. With this will come an incentive to refuse handouts, or insist upon a reward for my services. If I go with the second option, the scenario becomes (more or less) Ofek's "incendiary hub" scenario. If I go instead with my first option – simply refusing to bestow favors (and instructing family to do likewise) – we have a new scenario: call it the "fire island" scenario, on which good firekeepers share only with family and friends. There is no incentive for free-ridership in this scenario. But it is no market scenario, nor is it unstable – indeed it is simply the condition of all private goods. It promotes division of labor or specialization no more than does any other advantage conferred by consumption of a private good. And so, while free-ridership dooms the prospects of free "fire insurance", it does not doom the private production and consumption of fire.

What remains of Ofek's conclusion? It has to be qualified as follows: If fire does not remain privately produced and consumed (confined to its islands), it can become a catalyst of market exchange (and the division of labor that comes with it). On this story, fire does not realize the explanatory promise of "sparking" a new era.

However we believe that Ofek's reasoning is much more promising than this bland conclusion would suggest. Less is more: if we leave out the instability conjecture, we are left with a better argument. Compare the three scenarios on offer: "campfire", "private fire" and "incendiary hub". "Campfire" suffers from free-ridership, so it is inherently unstable. "Private fire" will either devolve into "fire island" or it will evolve into "incendiary hub." "Fire island" involves underutilization of fire (it's inefficient); "incendiary hub" will produce the right amount of fire utilization and is considerably more efficient. Nature will choose "incendiary hub". And we can perhaps reiterate one of Ofek's earlier remarks: If the players could not figure out their own interest for themselves in the short run (due to cognitive limitations), natural selection would figure it out for them in the long run.

But what is nature figuring out? *Nature is figuring out which is the better (i.e. more efficient) COMMUNITY organizational strategy vis-à-vis fire production and utilization*. On this amended reasoning, selection is taking place at a level considerably higher than the organism level. And it is specifically favoring a certain agency structure, as such.

We have nowise, by focusing upon contrived goods, avoided the need to address this issue. Any more than focusing upon public goods was able to do so. (Ofek, it's worth mentioning, is open to multiple levels of natural selection, so might view our amendment as reasonably friendly.) But how did the group-level entities become eligible to compete in the evolutionary dance in the first place? It becomes clear that whether you prefer to focus upon public goods or contrived goods, you do not avoid the need to answer this question. And without an answer to this question, the evolutionary story is incomplete. Our bonding story is an attempt to address that question.

VII. MORAL EMOTIONS? TRIBAL INSTINCTS?

We have suggested that bonding can help account for the prevalence of human cooperation that characterizes the organizational structures in which we live. The two existing types of models of cooperation that come closest to our own are (1) models that put moral emotions at center stage, and (2) models that put tribal instincts at center stage.

Frank: the moral-emotions model of cross-lineage cooperation

In his work on cooperation, Robert Frank (2001) argues that "moral emotions," such as sympathy, can account for human cooperation. He points to the fact that physical proximity and communication, which promote the development of sympathetic bonds, are conducive to cooperation in PD situations. For Frank, moral emotions support cooperation because they have subtle physiological effects that function as hard-to-fake signals that would-be cooperators can rely upon.

We have two main qualms with Frank's position. The first is that it is based on the hasty assumption that the only sort of bonding that accounts for human cooperation is emotional bonding. Physical proximity and communication can promote the development not only of emotional bonds, but also of identification; and, like emotional attachment, identification is conducive to cooperation in PD situations. Moreover, phenomenologically, identification seems like a more plausible candidate than emotional attachment when it comes to accounting for cooperation between individuals who are more or less strangers and whose contact has been limited to a brief, unintense encounter, like a ten-minute discussion session.

Our second qualm with Frank's position is that it casts the connection between emotional attachment and cooperation as less direct than it seems to be, given that emotional attachment is a form of bonding. For Frank, moral emotions support cooperation as follows: they give rise to hard-to-fake physiological manifestations of moral emotion, which serve as reliable, observable signals of genuine commitment, which support mutual trust, which supports mutual aid. This way of casting the connection between emotional attachment and cooperation suggests that if we could not signal our commitment to one another, there would be no mutual aid. But this seems wrong. If I am emotionally attached to you and you are emotionally attached to me (perhaps because, a few years ago, we went through a series of emotionally charged experiences together), this bond will support mutual aid, even if we cannot signal our commitment to one another. While mutual awareness of our mutual attachment may reinforce our bond to one another, the bond itself is sufficient to account for mutual aid; mutual awareness of the bond via signaling is not essential. Similarly, if I identify with you and you identify with me (perhaps because we have something in common), this bond will support mutual aid, even if we cannot signal our commitment to one another. This is evidenced by the fact (mentioned above) that individuals tend to exhibit solidarity with respect to anonymous individuals whom they think of as in the same boat as themselves. Note that, as the last few sentences suggest,

conditions that promote one's bonding to another will often promote this other's bonding to oneself as well; so the frequency of *mutual* aid, as opposed to one-sided commitment, is easily comprehensible, even without appeal to hard-to-fake signals.

Richerson and Boyd: the tribal-instincts model of cross-lineage cooperation

According to Peter Richerson and Robert Boyd, we humans have a tribal instinct that "allows us to interact cooperatively with … a rather large set of distantly related or unrelated individuals with culturally defined boundaries" (1999, 255-6). This tribal instinct prompts us to divide ourselves into groups in accordance with a variety of markers, and to demonstrate commitment to group goals. The relevant groups are tribe-like in that they are relatively egalitarian, with charismatic leaders exerting influence but not authoritarian control. Like Frank, Richerson and Boyd (2001) see emotional attachment as crucial to cooperation. They maintain that

> commitments to group goals are deeply rooted in the emotions of individual humans who make up the groups. The threats and promises of leaders are only credible to the extent that followers will collectively back them up with passionate action. (87)

But in Richerson and Boyd's model it is emotional attachment to the group, rather than to individuals in the group, that is crucial. Richerson and Boyd explicitly make room for progroup commitments that are not the result of personal attachments (2001, 187-8 & 202-3).

In defense of their tribal instincts model, Richerson and Boyd effectively argue that commitment to group goals and altruistic sacrifice are greater in military groups that have a tribal structure (or nested tribal structure) than in military groups that attempt to function via straightforwardly authoritarian control. We do not want to deny that humans may have a tribal instinct, but there are, we believe, cases of cooperation between individuals who are more or less strangers that do not fit neatly within Richerson and Boyd's tribal instincts model. Consider the following case, which Richerson and Boyd bring up in defense of their view that progroup commitment need not rest on personal attachments:

> In his prototypical experiments, Tajfel (1981) told subjects that they were participating in a test of aesthetic judgment. They were shown pictures of paintings by Klee and Kandinsky and asked to indicate which they preferred. Subjects were then divided into two groups, supposedly on the basis of their aesthetic preferences, but in fact at random. The subjects' task was then to divide a sum of money among members of their own group or the other group. Subjects discriminated in favor of the sham in-group members (2001, 203).

In this case, there is no charismatic leader and no basis for the attribution of emotional attachment to the group. (Unlike the members of the successful military groups Richerson and Boyd focus on, the group members in Tajfel's experiments do not have anything like a history with their group, during which an emotional bond to their group could develop; nor have they any emotionally charged experiences with their group.) At least on the face of it, the bonding in Tajfel's experiments seems like a clear case of simple identification.

The subjects seem to be conceiving of themselves as a part of a "we" and acting accordingly, just as one might conceive of oneself as temporally extended and act accordingly.

Of course, if one assumes that emotional attachment is the only form of bonding, then all bonds (including *intra*personal bonds) will be interpreted as emotional bonds. It will then be tempting to downplay the phenomenology of emotion. This is not necessary if one recognizes that bonding is a genus.

VIII. OUR CONTRIBUTION: THE NATURAL HISTORY OF BONDING SCHEMES

In the magnificent *Guns, Germs and Steel,* Jared Diamond seeks to explain variations in the fates of human societies entirely by means of features of their biogeographical and ecological realities. Everything else – including all of culture and all sources of technological innovation and accumulation – he treats as sources of noise. Selection retains only what is of advantage to the organism/population in the particular niche it happens to occupy or help construct. One construal of the proposal is that it gives voice to the idea that ecological realities, when they speak, speak decisively; ecology is always the highest and most important driving force when it comes to evolution. This "ecology is destiny" idea lies behind the approach to the problem of explaining the prevalence of human large-scale living as a purely ecological/dynamical problem. It lies at the heart of the generalist objection we sought to disarm at the outset.

And herein also lies its weakness. The "ecology is destiny" thesis is rarely true, and when true, true only in the limit, true therefore only as an approximation. For in every instance the organism is always part of the "ecology" to which it must adapt. This can matter a great deal. In some cases (as in the case of modern humans) the organism and its conspecifics create or construct conditions – niches – that are totally unlike those in which their species originally appeared, and that figure prominently in its future evolution. Furthermore, an organism can enter into alliances with conspecifics, and indeed with members of different species, that shield it from harsh "ecology", whether constructed or simply found. These bonds are forged serendipitously, perhaps. But once forged, natural selection honors them. Natural selection does not dictate that each individual organism must go it alone against its "ecology." If that were the case, then organisms as such – as alliances among cells, many genetically unrelated – should never have arisen.[22] Natural selection acts on such entities – such agencies – as it encounters. It does not arbitrarily cast asunder what has bonded together; and so what has bonded together might co-evolve. For a mammal, bonding *becomes* a feature of ecology (a feature of what one has to contend with), because it is part of what its conspecifics do; and it is at one and the same time also a feature of what comes naturally to it, the organism, as well. For humans, bonding comes in more varieties, because of the human ability to conceptualize, but the different varieties

22] This point draws on argumentation in Sober and Wilson (1998) as well as in Okasha (2009).

have the same motivational effects. Bonding with nonkin is thus motivationally no different from emotional bonding.

And this, fundamentally, is why a generalist approach is insufficient: there must be room made for analyzing what the organism brings to its niche. And so a purely dynamical/ecological analysis of an evolutionary problem is insufficiently fine grained. The organism changes the specifics of the ecological problem that it finds, however minutely. And the effect can be amplified by many orders of magnitude, even exponentially, when the organism can ally itself with others to change the question of *who* the ecological problem is a problem *for.*

To handle what requires handling, attention must be given to the nature of bonding schemes, their emergence, and their contributions to – and interference with – realities that antecede them. For this analysis an exhaustive taxonomy of bonding schemes is required, as well as mechanisms that can give rise to them. We have taken a first, baby step in this direction, challenging the supposed platitude that the boundaries of agencies coincide exactly, always and everywhere, with the contours of an organism's skin (or fur or what-have-you). We are contending that the assumption to the effect that strategy selection happens, always and everywhere, at the level of the individual – which is implicit in so much evolutionary game theoretical analyses – is without foundation.[23]

It is important for those interested in modeling choice situations (in game theory, for example) to acknowledge units of agentic organization.[24] So to acknowledge is also to acknowledge a new task for decision theory to perform: decision theory needs to address the question of how entities navigate between acting as individuals and acting as members of larger units. The thing to keep in mind is that the unit of agency can vary and a methodological decision to treat, invariably, only one possible unit of agency as *the* unit of agency can result in serious distortions. And it can result in loss of a capacity to explain other, choice-related phenomena.

By suggesting that certain large-scale social behavior of humans rests on a general capacity for bonding with others, which can manifest in a variety of ways, we are resisting the image of choosing with which Frederick Schick opens his book – and which we quote at the foot of the title of this essay: "Life is a long trip in a cheap car," he writes. "In a dark country. Without a good map." We reply that perhaps once upon a time, in primordial days, long before our own species appeared on this terrestrial stage, life might have been as Schick describes. And indeed it might be that way for many organisms on the planet today. But the human species has never known life this way. Life for the modern *homo*

23] All of the research conducted by Skyrms and colleagues falls under this category. Moreover, while Skyrms applies algorithms that mimic natural selection, he by no means models the mechanisms of reproduction that can be thought of as mammalian; indeed, it is best to say that he does not model reproduction at all – reproduction is treated exogenously, and not linked to any processes in the model whatsoever (the most telling fact is that the "young" do not differ in any way from the "adult" organisms). It is most charitable therefore to say that Skyrms has not yet ventured into the area of multigenerational solutions.

24] Susan Hurley (1989) similarly urges a non-fixed conception of the units of agency.

sapiens may be (if she's especially lucky) a long trip, but thanks to human bonding capacities, there is at least one well-lit road for every dark one, public transportation is the rule rather than the exception, and likely as not there is someone willing to go with you some of the way.

This is all good news. The hard-headed "realist" who insists that humans are motivated entirely by personal or individual concerns is, by our lights, no true realist. While humans may not often be *true* altruists, they are nonetheless prepared to put out for groups of which they conceptualize themselves as members, however they achieve this identification. It takes really quite little to achieve this conceptualization, as we have discussed. And this is good news indeed, since the fate of the globe now depends heavily upon human cooperation across multiple boundaries.

m.thalos@utah.edu
c.andreou@utah.edu

REFERENCES

Axelrod, Robert. 1984. *Evolution of Cooperation*. New York: Basic Books.

Axelrod, Robert and William Hamilton. 1988. The evolution of cooperation. *Science* 242: 1385-90.

Bacharach, Michael. 1999. Interactive team reasoning: A contribution to the theory of cooperation. *Research in Economics* 53: 117-47.

———. 2006. *Beyond Individual Choice*. ed. Natalie Gold and Robert Sugden. Princeton, New Jersey: Princeton University Press.

Barkow, J., L. Cosmides, and J. Tooby, ed. 1995. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Baron-Cohen, S. 1995. *Mindblindness*. Cambridge, MA: MIT/Bradford books.

Batterman, Robert. 1998. Game theoretic explanations and justice. *Philosophy of Science* 65: 76-102.

Binmore, Ken. 1994. *Playing Fair*. Cambridge, MA: MIT Press.

———. 1998. A utilitarian theory of political legitimacy. In *Economics, Values and Organization*, ed. Avner Ben-Ner and Louis Putterman. New York: Cambridge University Press.

Bolton, G. and A. Ockenfels. 2000. ERC – a theory of equity, reciprocity and competition. *American Economic Review* 90: 166-93.

Buss, David. 1994. *The Evolution of Desire: Strategies of Human Mating.* New York: Basic Books.

Buss, David and N. Malamuth, ed. 1996. *Sex, Power, Conflict: Evolutionary and Feminist Perspectives*. New York: Oxford University Press.

Byrne, Richard. 1995. *The Thinking Ape*. New York: Oxford University Press.

Byrne, R. W. and A. Whiten. 1985. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*. Oxford: Clarendon Press.

Carver, Charles and Michael Scheier. 2001. *On the Self-Regulation of Behavior*. New York: Cambridge University Press.

Cleckley, Hervey. 1941. *The Mask of Sanity*. Augusta,GA: Emily S. Cleckley.

Crawford, Charles and Dennis L. Krebs, ed. 1998. *Handbook of Evolutionary Psychology: Ideas, Issues and Applications.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Dawes, R. M., A. J. C. van de Kragt, and J. M. Orbell. 1997. Not me or thee but we: The importance of group identity in eliciting cooperation in dilemma situations: Experimental manipulations. In *Research on judgment and decision making*, ed. W. M. Goldstein and R. M. Hogarth. Cambridge: Cambridge University Press.

Demetriou, Andreas and Smaragda Kazi. 2001. *Unity and Modularity in the Mind and Self.* New York: Routledge.

Diamond, Jared. 1999. *Guns, Germs & Steel : The Fates of Human Societies*. London: W. W. Norton.

Duval, ThomasShelley, Paul J. Silvia and Neal Lalwani. 2001. *Self-awareness and Causal Attribution*. New York: Springer-Verlag.

Fehr, Ernst and K. M. Schmidt. 1999. A theory of fairness, competition and cooperation. *Quarterly Journal of Economics* 114: 817-68.

Geanakoplos, John, David Pearce, and Ennio Stacchetti 1989. Psychological games and sequential rationality. *Games and Economic Behavior* 1: 6-79.

Fisher, R. A. 1958. *The Genetical Theory of Natural Selection*. NY: Dover.

Frank, Robert. 2001. Cooperation through emotional commitment. In *Evolution and the Capacity for Commitment*, ed. Randolph M. Nesse. New York: Russell Sage Foundation.

Gold, Natalie and Robert Sugden. 2007. Theories of team agency. In *Rationality and Commitment*, ed. Fabienne Peter and Hans Bernhard Schmid. Oxford : Oxford University Press.

Hargreaves-Heap, Shaun and Yanis Varoufakis. 2004. *Game Theory: A Critical Text*. London and New York: Routledge.

Harris, Paul. 1991. The work of the imagination. In *Natural Theories of* Mind, ed. Andrew Whiten. Cambridge: Basil Blackwell.

———. 1994. Understanding pretence. In Charles *Children's Early Understanding of Mind* Lewis and Peter Mitchell. New Jersey: Lawrence Erlbaum and Associates.

———. 1995. From simulation to folk psychology: The case for development. In *Folk Psychology: The Theory of Mind Debate*, ed. Martin Davies and Tony Stone, 207-231. New York: Blackwell.

Hobbes, Thomas. 1994 [1668]. *Leviathan.* ed. Edwin Curley. Cambridge: Hackett Press.

Hrdy, Sarah Blaffer. 1999. *Mother Nature*. New York: Ballantine Books.

———. 2009. *Mothers and Others.* New Haven: Harvard University Press.

Hurley, Susan. 1989. *Natural Reasons*. Oxford University Press

Kramer, R. M. and M. B. Brewer. 1986. Social group identity and the emergence of cooperation in resource conservation dilemmas. *Experimental social dilemmas*, ed. H. Wilke, D. Messick, and C. Rutte. Frankfurt am Main: Verlag Peter Lang.

Maynard-Smith. 1982. *Evolution and the Theory of Games*. New York: Cambridge University Press.

Nagel, Thomas. 1970. *The Possibility of Altruism*. Princeton: Princeton University Press.

Nisbett, Richard. 2003. *Geography of Thought: How Asians and Westerners Think Differently...and Why.* New York City: Free Press.

Ofek, Haim. 2004. *Second Nature: Economic origins of human evolution*. Cambridge University Press.

Okasha, S. 2009. *Evolution and the Levels of Selection.* New York: Oxford University Press.

Panksepp, Jaak. 2003. Feeling the pain of social loss. *Science* 302 (5643): 237-239.

Rabin, Matthew. 1993. Incorporating fairness into game theory and economics. *American Economic Review* 83: 1281-302.

Richerson, Peter J. and Robert Boyd. 1985. *Culture and the Evolutionary Process.* University of Chicago Press.

———. 1999. Complex Societies: The evolutionary origins of a crude superorganism. *Human Nature* 10 (3), 253-89.
———. 2001. The evolution of subjective commitment to groups: A tribal instincts hypothesis. *Evolution and the Capacity for Commitment*, ed. Randolph M. Nesse. New York: Russell Sage Foundation.
Sally, D. 1995. Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society* 7: 58-92.
Schick, Frederick. 1997. *Making Choices*. New York: Cambridge University Press.
Selman, Robert. 1980. *The Growth of Interpersonal Understanding*. New York: Academic Press.
Skyrms, Brian. 1994. Sex and justice. *Journal of Philosophy* 91: 305-320.
———. 2003. *Stag Hunt and the Evolution of Social Structure*. New York: Cambridge University Press.
———. 1996. *Evolution of the Social Contract*. New York: Cambridge University Press.
Sober, E. and Wilson, D.S. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.
Sterelny, Kim. 2000. *The Evolution of Agency and Other Essays.* Cambridge: Cambridge University Press.
Strum, S.C., D. Forster and E. Hutchins. 1997. Why Machiavellian intelligence may not be Machiavellian. In *Machiavellian Intelligence II: Extensions and Evaluations*, ed. R. Byrne and A. Whiten. New York: Cambridge University Press.
Thalos, Mariam. 1999. Degrees of Freedom in the Social World. *Journal of Political Philosophy* 7: 453-77.
———. 2007. Sources of Behavior: Toward a Naturalistic, Control Metaphysics of Agency. *Distributed Cognition and the Will*, ed. Don Ross and David Spurrett, Harvard: MIT Press, 123-67.
———. 2008 . On Planning: Towards a Natural History of the Will. *Philosophical Papers* 37: 289-317.
Tomasello, Michael. 1999. *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press.
Wegner, Daniel and Robin Vallacher, 1980, eds. *Self in Social Psychology*. New York: Oxford University Press.
Wilson, E. O. 1975. *Sociobiology: The New Synthesis*. Cambridge: Harvard University Press.
Wilson, E.O. and Bert Hölldobler. 1990. *The Ants.* Cambridge, Mass: Harvard University Press.
———. 2009. *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies.* New York: W.W. Norton & Company.