

The Social and Mental Dynamics of Cooperation

SOCCOP

Section A – Collaborative Research Project (CRP)

1. Description of the CRP

a. International state-of-the-art of the evolutionary theory of cooperation

Despite the fact that human cooperation has been a central concern of biology and the human behavioural sciences since Adam Smith and Charles Darwin, there is currently no accepted model of how the cognitive capacities and motivational predispositions of humans interact to support cooperation and how these unique human traits evolved. Contemporary evidence from neuroscience, behavioural genetics, and behavioural game theory suggests that while the behavioural disciplines (biology, economics, anthropology, psychology, and sociology) have made major contributions to understanding cooperation, each discipline ignores a key part of the overall picture. The nature of cooperation is such that it does not break down into independent parts susceptible to isolated analysis by specific disciplines. The *SOCCOP* project is a novel and uncompromisingly trans-disciplinary strategy for understanding human cooperation. We anticipate a number of specific scientific breakthroughs that promise to have practical application to problems of organizational governance, conflict resolution, and the promotion of intergroup tolerance.

b. Aims and objectives

Simply bringing researchers together from different disciplines to work jointly will not suffice to meet the challenge of understanding cooperation. This because the various disciplines have incompatible core models of human behaviour (Gintis, 2006). However, there has been considerable progress in overcoming the disciplinary hurdles in the past decade through an intense dialogue between empirical and theoretical studies using such transdisciplinary tools such as behavioural and evolutionary game theory, agent-based modelling, and neuroscience. This experience is the backdrop of Gintis' suggested framework for the *unification of behavioural research*. Our TECT proposal is the first comprehensive transdisciplinary research project fully informed by this framework. Relying on past experience and an innovative methodology, we expect a degree of cross-disciplinary coordination and interaction well beyond anything that has hitherto proven possible.

Specifically, SOCCOP aims to:

- (a) develop consistent and stable behavioural measures that can be used in other studies;
- (b) perform behavioural studies on a group of >600 individuals;
- (c) evaluate the empirical plausibility of distinct models of the evolution of human cooperation;
- (d) combine game theory and behavioural genetics to understand the genetic basis of phenotypic behavioural heterogeneity in humans;
- (e) locate genetic correlates of behaviour in strategic interaction, with behavioural evidence from behavioural game theory;
- (f) identify both the pro-social and anti-social aspects of human cooperative tendencies;
- (g) develop a new approach to organizational governance, public policy and economic incentives taking into account the cooperative nature of humans;
- (h) publish results in high-impact journals and international meetings.

c. Strategy and work plan

SOCCOP's research strategy rests on shared theoretical assumptions which establish the common framework of the proposed research activities. These assumptions are the following:

- (1) Human society is the product of an ***evolutionary dynamic***. While our species is by far the most advanced among the variety of social species to have evolved, it is subject to the same deep evolutionary principles. This means that the proximate mechanisms of human behaviour are likely to be the product of ultimate evolutionary mechanisms, and are likely to be revealed by a thorough analysis of both the continuity and the rupture of human behaviour with that of other species.
- (2) Human society is a ***complex adaptive system***. This means that it is likely to be understood not by one overarching model, but by a variety of interrelated models, some highly analytical and mathematical, others agent-based, and still others historical and ethnographic.
- (3) Human society is the product of ***gene-culture coevolution***, culture not only being the *differentia specifica* of humanity, but also the central environmental dimension accounting for the unique ensemble of genetically-based capacities for cognition, communication, emotion, and even aesthetics.
- (4) In place of the rational actor model, we will adopt the more general ***beliefs, preferences, and constraints*** (BPC) approach (see Gintis IP and Gintis 2006), both because of the centrality of the socially produced nature of beliefs and preferences (overlooked in classical game theory), and because the term "rational" is irremediably saddled with misleading connotations to which we do not ascribe.

It is worth noting that the highly ***trans-disciplinary*** nature of the TECT project demands more than the usual attention to fostering collaborations and ensuring proper dissemination. It is a key objective of the both European Science Foundation and the U.S. National Science Foundation to enhance cross

national collaboration. To meet these priorities, in addition to pursuing the research agenda discussed here and reporting the results in various high-level academic publications, we will also support networking and cooperative research activities of the project by appointing post-doctoral researchers acting as intermediaries among various research groups. Researchers will be sponsored for short or extended visits to project sites. At least three meetings (kick-off, mid-term, final) of the researchers involved at the project will be necessary to evaluate research progress and to consider methods for dissemination.

d. Bibliography

Gintis, Herbert, "A Framework for the Unification of the Behavioural Sciences," *Behavioural and Brain Sciences (In Press)* (2006).

_____, Samuel Bowles, Robert Boyd, and Ernst Fehr, *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life* (Cambridge: The MIT Press, 2005).

Kerszberg, Michel, "Genetics and Epigenetics of Neural Function: a Model," *Journal of Cognitive Neuroscience*. 2:51-7 (1990).

Navarro, A. and N. H. Barton, "Chromosomal speciation and molecular divergence. Accelerated evolution of genes in rearranged chromosomes." *Science* 300: 321-324 (2003).

Számadó, Szabolcs & Szathmáry, Eörs: "Competing selective scenarios for the emergence of natural language," *Trends Ecol Evol*, in press (2006).

e. Deliverables and milestones

Joint deliverables include:

- three meetings of all participants in the project (kick-off, mid-term, final);
- multilateral meetings of PIs throughout the project's duration to enhance interaction within and among research groups;
- publications in high-impact journals and presentation of results at international meetings.

Deliverables assigned to one or more PIs include:

Samuel Bowles (Santa Fe Institute)

- hosting the mid-term meeting of SOCCOP researchers in 2008-9 at the Santa Fe Institute to evaluate research progress and to consider methods for its dissemination;

- appointment of post-doctoral (in exceptional cases, pre-doctoral) researchers sharing time between the Santa Fe Institute and European labs/research sites;
- sponsoring researchers for short or extended visits to the Santa Fe Institute and other project sites.

Institutional Niche Construction

- a series of papers including:

one paper on the ecological and social conditions under which cooperation is sustained in behavioural field experiments;

two papers exploring how learning rules may co-evolve with cooperative behaviours (with Rowthorn and Sickert);

one paper on the distinctive nature of reproductive levelling in humans hunter gatherer populations by comparison with other primates;

one or more papers performing empirical tests of the reproductive levelling and other forms of niche construction as a contributor to the evolution of genetic and cultural dispositions to altruistic cooperative behaviours (possibly in collaboration with Fehr and Navarro);

an empirical and theoretical paper on the emergence of the distinctive forms of (private) property.

Governing the cooperative species

- a series of papers including:

one paper on a model of potentially perverse effects of explicit incentives (and also for the opposite, termed 'crowding in') with experimental data and historical case studies.

one paper reformulating the conventional approach to optimal incentives.

- non-technical introductory book (Bowles).

Proximate explanations for human and animal cooperation (Cristiano Castelfranchi)

- experiments with the non-human primates and with humans (Year 1);

- publication of all preliminary results on the three objectives (theoretical, mathematical and computational models) (Year 1);

- interdisciplinary conference to present results and foster interaction with other CRPs (Year 2);

- book on theoretical, computational, mathematical and experimental results of the IP (Year 3).

Ernst Fehr

- establish stable phenotypes and accumulate relevant data;

- present results within three years;

- top-notch publications in leading journals.

Herbert Gintis

- book on the topic of unification of the behavioural sciences: to include results from SOCCOP (with other PI's as chapter authors where appropriate).

Michel Kerszberg and Eörs Szathmáry

- publication of research papers in high-ranked journals;
- synthetic summary of the results of their work, possibly in book form.

Ruth Mace

- piloting fieldwork methods (Year 1)
- full experimental protocol (Year 1)
- database on ecological and individual variation in behaviour in economic games (Year 2);
- completing analysis on individual and group level determinants of prosocial norms (Year 3);
- assessing assumptions of various models of co-operative norms in the light of empirical findings (Year 3).

Arcadi Navarro

- developing measures and performing laboratory experiments (Year 1&2);
- genotyping (Year 2&3);
- data analysis (Year 3) of the third year;
- functional and comparative studies (Year 2&3).

f. Total overall budget

See below:

A.2. Description of the collaboration

The Principal Investigators (PIs) have been invited with a view to ensuring the ***multidisciplinarity of the CRP***. PIs and their Associates have long-standing and on-going research interests that they believe will be enhanced by sustained direct contact with one another. All are deeply interested in and informed concerning the work of the others, and believe that innovative results will emerge from the collaboration that flows from frequent contact and intellectual exchange.

The Individual Projects (IPs) summarized below rely on both theoretical and empirical methods. The research objectives of the IPs focus on the fundamental structural, interactional and mental determinants of the evolution of cooperation among unrelated individuals in humans and other species. The IPs will complement one another insofar as they address ***the four major areas in the study of the evolution of cooperation***: (i) cognitive and linguistic capacities required for and exercised in exchange and cooperation; (ii) the nature of other-regarding preferences and the linguistic and cognitive capacities that support such preferences; (iii) nature of the individual behaviours and social institutions that induce self-regarding agents to cooperate, and (iv) the biological and social conditions for the evolution of cooperative behaviours. Therefore, the proposed collaboration promises unique results as it combines self-interested and social preference approaches to understanding cooperation and can thus provide an analysis of both the proximate causes of cooperation and its evolutionary origins. The collaborative effort will centre on the following research topics:

The Cooperative Mind

This topic incorporates several interrelated questions. First, what are the cognitive and emotional determinants of widespread cooperation in humans? Second, how have distinctive human institutions facilitated the evolution of altruism and cooperation among early humans, by means of reproductive levelling and parochial, in-group altruism? Third, how do human decisions concerning cooperation differ from those of other primates? What are the cognitive and emotional prerequisites for widespread cooperation in humans? Fourth, do properties associated with social organization (e.g. distribution and accessibility of resources, degree of immigration/emigration, population size) affect the norms adopted by a group? Fifth, what are the processes that drive the cultural evolution of prosocial behaviour? What are the time scales over which they occur? Sixth, how does group size evolve as an endogenous variable in human groups? How does migration affect the structure and pace of gene-culture coevolution? These questions will be addressed by Bowles, Castelfranchi, Mace, and Rowthorn. Further questions are: Seventh, what is the relationship between the capacity of a society to punish normative infractions, especially dishonesty, and the evolution of language (Kerszberg, Szathmary, Gintis)? And finally, eighth, what is the role of emotion vs. cognition and of self-interest vs. altruism

in the efficacy of gossip to create the public information on which coordinated group reward and punishment are predicated (Castelfranchi, Gintis)?

The Genetic Basis of Behavioural Heterogeneity

Modern humans are the only known species to show extraordinary variation in individual behaviour, important aspects of which have been linked by behavioural geneticists to genetic heterogeneity. Such heterogeneity indicates either genetic/cultural disequilibrium or frequency dependence, and the former is unlikely to have persisted as long as it has been observed and documented. Frequency dependence, the more likely cause is, by definition, in the realm of game theory. Surprisingly, the intimate connection between game theory and behavioural heterogeneity has not been exploited by either behavioural geneticists or game theorists. We aim to correct this.

Behavioural experiments performed in this area will be designed in coordination with the ones of Bowles and Mace described above. Variability in some complex behavioural traits related to cooperation can now be accurately measured by means of the protocols of experimental economics (Fehr and Gächter, 2002). Two Principal Investigators and an Associate will jointly address this issue. Alex Kacelnik will undertake a theoretical and empirical analysis of individual and group decision-making under risk. Ernst Fehr and his co-workers will undertake an ambitious empirical study of the behavioural genetics of risk, time preference, and social preferences, including reciprocity and inequality aversion, funded by another agency. Navarro, in coordinate with Fehr and Kacelnik, will explore the behavioural genetics of trade and cooperation.

Self-regarding Cooperation

The conditions under which self-interested individuals cooperate have been explored using evolutionary models of two person interactions and static, classical game theoretical, models of cooperation among large numbers of agents (e.g. the Folk Theorem in repeated game theory). Bowles, Gintis, and Rowthorn will fill a gap in this literature by exploring the empirical plausibility and dynamic properties of self-interest based models with two realistic characteristics: large numbers of participants and significant amounts of noise and private information. Our hypothesis is that the self-interest based models may not provide an adequate explanation of the evolution of cooperation among large numbers of unrelated individuals except under empirically implausible assumptions concerning time discounting and the accuracy and availability of information among potential cooperators.

Institutional Niche Construction

Reproductive levelling behaviours such as sharing information and food and monogamous mating systems reduce differences in fitness and material success within groups. Bowles, Rowthorn, and their co-workers will address the central gene-culture coevolutionary fact that socialization institutions bias

developmental processes towards group beneficial cooperative behaviours, and hence bias genetic evolution towards prosociality. Using data on late Pleistocene and early Holocene conditions, as estimated from the literature and from the results of behavioural heterogeneity, we will explore whether these aspects of social structure could have co-evolved with individual social preferences that would have been unviable in the absence of these institutional niches. We hypothesize that this may have occurred through inter-demic selection: groups with these constructed niches would tend to have more cooperative individuals and in turn would therefore be more likely to survive environmental and military challenges.

Parochial Altruism and its Evolution

Many animals recognize and condition their behaviour on group membership, often, as with chimpanzees, ants and humans, favouring insiders and inflicting lethal costs on outsiders. Insider bias may support within group cooperation, but it also impedes mutually beneficial exchange with outsiders. We will address this puzzle in the case of humans. Bowles, Gintis, Rowthorn and their co-workers will address the hypothesis that parochial and altruistic individuals evolved because groups in which they were prevalent had a greater probability of survival during periods of intergroup conflict, which we will show was probably common during the late Pleistocene and early Holocene.

Applications to Social Policy and Institutional Formation

A key goal of our project is to apply our research to social policy. In the final stages of our research, several of the Principal Investigators, including Bowles, Gintis, Fehr, and Rowthorn, will study how an understanding of human capacities for cooperation may contribute to the design of more effective public policies, high performance firms and other institutions. Evidence from experiments and historical case studies suggest that policies and institutions that are designed to work well if individuals are entirely self-regarding will not generally be the best when a significant fraction of the population is other-regarding. We will use behavioural experiments and models of complex social interactions to explore our hypothesis that policies designed to mobilize self-interest for the public good may crowd out social preferences, resulting in inferior outcomes.

Section B – Individual Projects (IPs)**IP 1: *A Framework for the Unification the Behavioural Sciences******Herbert Gintis******Central European University*****B.1. Individual Project contribution to the CRP****a. IP aims and objectives**

The various behavioural disciplines, including economics, biology, anthropology, sociology, psychology, and political science, model human behaviour in distinct and incompatible ways. Among the major incompatibilities is the treatment of the individual as rational and self-interested in economics and biology, and the treatment of human social institutions as mechanisms fostering efficient cooperation among self-interested individuals. This contrasts sharply with the view of the individual in social psychology, sociology, and anthropology as a repository of cultural values, and the treatment of cooperation in society as a product of the internalization of cultural norms. Another major incompatibility, cutting across disciplinary lines in the same way, is the use of game theory and decision theory in economics and biology to explain human behaviour, and the virtually complete rejection of these tools in the other disciplines. These incompatibilities show that the behavioural sciences fail to attain the status of true science, not because discrepancies exist, but because behavioural scientists have failed to address and adjudicate their differences. This causes severe problems for governmental scientific research funding agencies, whose behavioural budgets tend to be constrained by the unwillingness of policy makers and the tax-paying public to underwrite research activities with glaringly incompatible presumptions, yet whose principal investigators have not intention of confronting the thorny issues of interdisciplinary incompatibility.

b. Methodologies/experiments

The behavioural sciences all include models of individual human behaviour. Therefore, these models should be compatible, and indeed, there should be a common underlying model, enriched in different ways to meet the particular needs of each discipline. Realizing this goal at present cannot be easily attained, since the various behavioural disciplines currently have *incompatible* models. Yet, recent theoretical and empirical developments have created the conditions for rendering coherent the areas of overlap of the various behavioural disciplines, as outlined in this paper. The analytical tools deployed in this task incorporate core principles from several behavioural disciplines.

The standard justification for the fragmentation of the behavioural disciplines is that each has a model of human behaviour well suited to its particular object of study. However, where these objects of study *overlap*, their models must be compatible. In particular, psychology, economics, anthropology, biology, and sociology should have concordant explanations of law-abiding behaviour, charitable giving, political corruption, and voting behaviour, and other complex behaviours that do not fit nicely within disciplinary boundaries. They do not.

When we assumed directorship of a transdisciplinary research group aimed at modelling human decision-making and strategic interaction some twelve years, Robert Boyd (Anthropology, UCLA) and I confronted these and other incompatibilities directly, and worked out a research plan for overcoming them. The considerable experience I gained from interacting with the members of this research group deepened my understanding of the requirements for a unified behavioural science, and I have been working on a set of analytical principles whose application will eliminate interdisciplinary contradictions and permit a more seamless integration of interdisciplinary research. I currently have a target article in *Behavioural and Brain Sciences*, with comments by thirty-one behavioural scientists. I propose to draw on the research experience of the TECT project to further deepen and test the framework I have developed. During the course of the grant, I will prepare a book on the topic of the unification of the behavioural sciences, with contributions by several authors in specialized fields, including neuroscience, anthropology, and cognitive psychology.

Perhaps there was a time when this cross-disciplinary incoherence could be tolerated. In the past, economics dealt with growth, income distribution, industrial regulation and the business cycle, sociology dealt with culture, social stratification, and deviance, and psychology dealt with mental illness, processing of sensory inputs, and learning. Today, pressing social policy issues often fall squarely in the overlap of the behavioural disciplines, concerning such issues as the structure of the family, drug addiction, crime, corruption, tax compliance, social inequality, racial and ethnic tolerance and discrimination, and religious conflict. No longer can this incoherence be tolerated.

c. Work plan

The following are the main points in my framework for unifying the behavioural sciences. Two major categories, evolution and game theory, cover *ultimate* and *proximate* causality. Under each are subcategories that relate to overlapping interests of two or more behavioural disciplines. I will argue the following points:

1. Evolutionary Perspective: Evolutionary biology underlies all behavioural disciplines because *Homo sapiens* is an evolved species whose characteristics are the product of its particular evolutionary history. This idea is of course expressed by all of the IP's, especially Bowles, Castelfranchi, Kerszberg, and Szathmary.

1a. Society as a Complex Adaptive System: Understanding human society requires a combination of analytical model-building and testing, agents-based modelling and testing, and such non-analytical scientific endeavours as ethnographic, cultural, and biographical research. To capture correctly the interplay of analytical and synthetic studies, we must recognize that human society is, quite literally, a *complex adaptive system*. A complex system consists of a large population of similar entities (in our case, human individuals) who interact through regularized channels (e.g., networks, markets, social institutions) with significant stochastic elements, without a system of centralized organization and control (i.e., if there a state, it controls only a small fraction of all social interactions, and itself is a complex system). We say a complex system is *adaptive* if it evolves through some evolutionary (e.g., genetic, cultural, agent-based silicon) process of hereditary reproduction, mutation, and selection (Holland, 1975). To characterize a system as complex adaptive does not *explain* its operation, and does not *solve* any problems. However, it suggests that a wide variety of research tools, often considered as antithetical to one another, are likely to be jointly effective. Such novel research tools are needed because a complex adaptive system generally has *emergent properties* that cannot be analytically derived from its component parts (Morowitz, 2004). The complex adaptive nature of human society explains why the PI's are concerned with agent-based models of cooperation (Bowles, Castelfranchi and communication, and why anthropological evidence is prominent in many accounts (Bowles, Mace)

1b Gene-culture Coevolution: The centrality of culture and complex social organization to the evolutionary success of *Homo sapiens* implies that individual fitness in humans will depend on the structure of cultural life. Since obviously culture is influenced by human genetic propensities, it follows that human cognitive, affective, and moral capacities are the product of a unique dynamic known as *gene-culture coevolution*. This coevolutionary process has endowed us with preferences that go beyond the self-regarding concerns emphasized in traditional economic and biological theory, and embrace such other-regarding values as a taste for cooperation, fairness, and retribution, the capacity to empathize, and the ability to value such constitutive behaviours as honesty, hard work, toleration of diversity, and loyalty to one's reference group. Gene-culture coevolution is prominent in the work of Bowles, Castelfranchi, Kerszberg, and Szathmary.

1c. Imitation and Conformist Transmission: Cultural transmission generally takes the form of *conformism*: individuals accept the dominant cultural forms, ostensibly because it is fitness-enhancing

to do so (Bandura, 1977, Boyd and Richerson, 1985, Conlisk, 1988, Krueger and Funder, 2004). While adopting the beliefs, techniques, and cultural practices of successful individuals is a major mechanism of cultural transmission, there is constant cultural mutation, and individuals may adopt new cultural forms when they appear to better serve their interests (Gintis 1972, 2003a; Henrich 2001). One might expect that the analytical apparatus for understanding cultural transmission, including the evolution, diffusion, and extinction of cultural forms, might come from sociology or anthropology, the disciplines that focus on cultural life, but such is not the case. Both fields treat culture in a static manner that belies its dynamic and evolutionary character. By recognizing the common nature of genes and culture as forms of information that are transmitted intergenerationally, biology offers an accurate analytical basis for understanding cultural transmission.

1d. Internalization of Norms: In sharp contrast with other species, human preferences are *socially programmable*. Culture thus takes the form not only of information allowing superior control over nature, but also of *norms and values* that are incorporated into individual preference functions through the sociological mechanism known as *socialization* and the psychological mechanism known as the *internalization of norms*. Surprisingly, the internalization of norms, which is perhaps the most singularly characteristic feature of the human mind, and central to understanding cooperation and conflict in human society, is ignored or misrepresented in the other behavioural disciplines, anthropology aside.

2. Game Theory: The analysis of living systems includes one concept that does not occur in the non-living world, and is not analytically represented in the natural sciences. This is the notion of a *strategic interaction*, in which the behaviour of agents is derived by assuming that each is choosing a *fitness-relevant response* to the actions of other agents. The study of systems in which agents choose fitness-relevant responses and in which such responses evolve dynamically, is called *evolutionary game theory*. Game theory provides a transdisciplinary conceptual basis for analyzing choice in the presence of strategic interaction. However, the classical game theoretic assumption that agents are self-regarding must be abandoned except in specific situations (e.g. anonymous market interactions), and many characteristics that classical game theorists have considered deductions from the principles of rational behaviour, including the use of backward induction, are in fact not implied by rationality. Evolutionary game theory, whose equilibrium concept is that of a stable stationary point of a dynamical system, must thus replace classical game theory, which erroneously favours subgame perfection and sequentiality as equilibrium concepts. I will interact with the other PI's to bring out the game-theoretic casting of their projects and results.

2a. The Brain as a Decision Making Organ: In any organism with a central nervous system, the brain evolved because centralized information processing entailed enhanced decision making capacity,

the fitness benefits more than offsetting its metabolic and other costs. Therefore, decision making must be the central organizing principle of psychology. This is not to say that learning (the focus of behavioural psychology) and information processing (the focus of cognitive psychology) are not of supreme importance, but rather that principles of learning and information processing only make sense in the context of the decision making role of the brain.

2b. The Rational Actor Model: General evolutionary principles suggest that individual decision making can be modelled as optimizing a preference function subject to informational and material constraints. Natural selection leads the content of preferences to reflect biological fitness. The principle of expected utility extends this optimization to stochastic outcomes. The resulting model is called the *rational actor model* in economics, but I will generally refer to this as the *beliefs, preferences, and constraints* (BPC) model to avoid the often misleading connotations attached to the term "rational."

While accepting the above framework may entail substantive reworking of basic theory in a particular discipline, I expect that much research will be relatively unaffected by this reworking. For instance, a psychologist working on visual processing, or an economist working on futures markets, or an anthropologist tracking food sharing practices across social groups, or a sociologist gauging the effect of dual parenting on children's educational attainment, might gain little from knowing that a unified model of decision making underlay all the behavioural disciplines. But, I suggest that in such critical areas as the relationship between corruption and economic growth, community organization and substance abuse, taxation and public support for the welfare state, and the dynamics of criminality, researchers in one discipline are likely to benefit greatly from interacting with sister disciplines in developing valid and useful models.

d. Deliverables/milestones

I will produce a book on the above topic by the end of the project, including other PI's as chapter authors where appropriate.

e. Works cited

Bandura, Albert, *Social Learning Theory* (Englewood Cliffs, NJ: Prentice Hall, 1977).

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

Conlisk, John, "Optimization Cost," *Journal of Economic Behaviour and Organization* 9 (1988):213-228.

Gintis, Herbert, "A Radical Analysis of Welfare Economics and Individual Development," *Quarterly Journal of Economics* 86,4 (November 1972):572-599.

_____, "The Hitchhiker's Guide to Altruism: Genes, Culture, and the Internalization of Norms," *Journal of Theoretical Biology* 220,4 (2003):407-418.

Henrich, Joseph, "Cultural Transmission and the Diffusion of Innovations," *American Anthropologist* 103 (2001):992-1013.

Holland, John H., *Adaptation in Natural and Artificial Systems* (Ann Arbor: University of Michigan Press, 1975).

Krueger, Joachim I. and David C. Funder, "Towards a balanced social psychology: Causes, Consequences, and Cures for the Problem-seeking Approach to Social Behaviour and Cognition," *Behavioural and Brain Sciences* 27,3 (June 2004):313-327.

Morowitz, Harold, *The Emergence of Everything: How the World Became Complex* (Oxford: Oxford University Press, 2004).

Wrong, Dennis H., "The Oversocialized Conception of Man in Modern Sociology," *American Sociological Review* 26 (April 1961):183-193.

IP2: Institutional niche construction and the evolution of a cooperative species

Samuel Bowles

Director, Behavioural Sciences Program, Santa Fe Institute

B.1. Individual Project contribution to the CRP

a. IP aims and objectives

The Santa Fe Institute work group will contribute to two aspects of the project's research: (a) explaining how distinctive human institutions (within group reproductive levelling and between group lethal competition) may have facilitated the evolution of (within group) altruism and cooperation among early humans as well as outgroup hostility, and (b) exploring the implications of recent advances in behavioural economics and cognitive psychology for improving the governance of cooperative and trading relationships in firms, neighbourhoods and nations. The group will be headed by Samuel Bowles and will include Ruth Mace (anthropology, UCL), Robert Rowthorn (economics, Cambridge), Jung-Kyoo Choi (economics, Kyoongpook National University, Daegu) and Carlos Sickert (sociology, Universidad Catolica de Santiago), as well as collaborations with the TECT project's research teams headed by Cristiano Castelfranchi (cognitive sciences, CNR, Rome), Eranst Fehr (economics, Zurich), Herbert Gintis (economics CEU), and Arcadi Navarro (genetics, Pompeu Fabra, Barcelona).

b. Methodologies/experiments

Institutional niche construction

This project will use evolutionary game theory and agent-based simulations as well as empirical sources to model two of the ways that humans have created niches allowing for the evolution of social preferences. The first are strong insider outsider distinctions and frequent intergroup conflict, and the second are systems of within group reproductive levelling by means of resource sharing (and related variance reduction practices.). Taken together these institutions are represent a constructed niche facilitating the evolution of the social preferences supporting within-group cooperative behaviours. The analytical methods will follow the gene-culture coevolution framework outlined in the proposal by Gintis.

These models will be subjected to empirical tests for plausibility using what is known from archaeological, climatic, genetic, linguistic and other data about conditions during the late Pleistocene.

As institutional niche construction provides a possible basis for the evolution of a genetic predisposition to behave cooperatively, this work group will closely complement and collaborate with the work of the Navarro work group in Barcelona and the Fehr group in Zurich. The approach will model the co-evolution of group-level traits (the institutions in question) and individual traits (altruistic punishment and other social preferences). The work is important because while the essential role of these institutions in human evolutionary processes is recognized, their emergence, persistence and proliferation is not well understood.

Evolution of the cooperative species

This work group will explore the implications of the cooperative nature of our species for the design of public policies and institutions. The proposed research is important because policies and institutions that are designed to work well if citizens and economic actors are entirely self interested will not generally be the best institutions for a heterogeneous population in which significant numbers are motivated by strong reciprocity motives, inequality aversion, or other social preferences.

c. Work plan

Institutional niche construction

Background: Among the unique human capabilities – as is evident in the proposals of the work groups headed by Szathmary-Kerszberg and Castelfranchi make clear – are the linguistic and cognitive capacities that allow us to devise and secure generalized conformity to the complex institutions that induce high levels of cooperation. This uniquely human group-level institution-building capacity is central to the explanation of cooperative behaviours.

Reproductive levelling: Resource and information sharing, monogamous mating, possession-based property rights and other within group-variance reducing and competition attenuating institutions (Smith and Szathmary (1995)) alter evolutionary processes by reducing and perhaps even reversing the force of within group selection against altruistic behaviours. This project will devise methods of modelling the evolutionary dynamics of the various forms of reproductive levelling and assessing their likely empirical importance.

Parochial altruism: The hypothesis motivating this aspect of the work is that within-group altruism and between group parochialism (or aggression) co-evolved. We will model the evolution of a within group altruistic trait and a between group aggressiveness trait under empirical conditions approximating the late Pleistocene. The project will include empirical aspects (assessing the archaeological and other evidence for the likely importance of group conflict during the Pleistocene) as well as both analytical and simulation-based modelling.

Previous research: Since 1998 the Santa Fe Institute has hosted an on going working group (convened by Bowles, with Boyd, Young, Blume and Gintis as core members) on the co-evolution of individual behaviours and group-level institutions. The work most closely related to the insider-outsider parochialism project is McElreath, Boyd, and Richerson (2003) though this paper does not explain the endogenous emergence of hostility (and warfare) towards out-groups. The emergence and persistence of institutions has been studied using stochastic evolutionary game theory by Young, Kandori Mailath and Rob, and others, but these models have not been applied to question of variance- and competition-reduction or to the evolution of early humans. The contribution of warfare to the evolution of human cooperation has been suggested by Darwin, Boehm, Eibl-Eibesfeld and others but the relevant dynamics have not been studied.

Evolution of the cooperative species

Background: This project will reconsider and suggest alternatives to the conventional approach to policy making and constitution building initiated by Thomas Hobbes and John Locke, and embraced by many modern day economists, policy makers, and legal professionals. Contemporary versions of this approach seek to determine the contracts, property rights, and rules B in short, constitutions B that can lead self-regarding individuals to interact in such a way that the aggregate outcome is socially desirable. There is now considerable evidence that the kinds of incentive-based policies generally favoured by economists sometimes degrade performance rather than improve it. For example, when crèches in Haifa imposed a fine for parents arriving late to collect their children, lateness increased. (Gneezy and Rustichini (2000)).

Objectives: The proposed project will accomplish two things. It will re-analyze experimental evidence to isolate effect of social preferences on optimal policy and institutional design, and extending a standard model of optimal incentives explore the implications for optimal taxation, deterrence, and work motivation.. Among the hypotheses to be explored are the following. First, incentives that are well designed to induce self-interested individuals to implement a socially desired outcome may Achange the subjects frame from ethical to income-maximizing@ (Hauser, Xiao, McCabe, et al. (2004)) and thus reduce the salience of social preferences, leading to an inferior outcome. Second, even a relatively small number of those with social preferences (strong reciprocators, for example, who are willing to engage in altruistic punishment of defectors) can substantially alter equilibrium outcomes.

Prior research: The suggestion that the use of explicit incentives may backfire was initially made in Taylor developed by Frey and suggested in some empirical works (e.g. Somanathan (1991)). But these works did not attempt a reformulation of the canonical model of policy making and optimal incentives embodied in implementation theory.

d. Deliverables/milestones

Institutional niche construction

- A series of papers including:

one paper on the ecological and social conditions under which cooperation is sustained in behavioural field experiments;

two papers exploring how learning rules may co-evolve with cooperative behaviours (with Rowthorn and Sickert);

one paper on the distinctive nature of reproductive levelling in humans hunter gatherer populations by comparison with other primates;

one or more papers performing empirical tests of the reproductive levelling and other forms of niche construction as a contributor to the evolution of genetic and cultural dispositions to altruistic cooperative behaviours (possibly in collaboration with Fehr and Navarro);

an empirical and theoretical paper on the emergence of the distinctive forms of (private) property.

Evolution of the cooperative species

- A series of papers including:

one paper on a model that accounts for the sometimes perverse effects of explicit incentives (and also for the opposite, termed ‘crowding in’), illustrating the model’s insights by means of existing experimental data and historical case studies;

one paper reformulating the conventional approach to optimal incentives;

a short non-technical book written for both academic and policy engaged readers will be considered (Bowles).

- Hosting the mid-term meeting of SOCCOP researchers in 2008-9 at the Santa Fe Institute to evaluate research progress and to consider methods for its dissemination;

- Appointment of post-doctoral (in exceptional cases, pre-doctoral) researchers sharing time between the Santa Fe Institute and European labs/research sites;

- Sponsoring researchers for short or extended visits to the Santa Fe Institute and other project sites.

e. Justification for budget items

The highly trans-disciplinary nature of the TECT project demands more than the usual attention to fostering collaborations and ensuring proper dissemination. Because an objective of the both European Science Foundation and the U.S. National Science Foundation is to enhance cross national collaboration, the Institute will support the networking and cooperative research activities of the

project in the following ways. "At large" post-doctoral (and in exceptional cases, pre doctoral) researchers will be appointed with the expectation that they will divide their time between residence at the Institute and at one or more European labs (or other research sites) of the project. Researchers will be sponsored for short or extended visits to the Institute and to other project sites. Finally, a meeting of researchers involved at the project will be held in 2008-9 at the Institute to evaluate research progress and to consider methods for its dissemination. The all-researcher meetings are designed to synthesize the research of the various collaborating components of the project.

f. Works cited

Gneezy, Uri and Aldo Rustichini. 2000. "A Fine is a Price." *Journal of Legal Studies*, 29:1, pp. 1-17.

McElreath, Richard, Robert Boyd, and Peter Richerson. 2003. "Shared Norms and the Evolution of Ethnic Markers." *Current Anthropology*, 44:1, pp. 122-29.

Smith, J. Maynard and Eors Szathmary. 1995. *The Major Transitions in Evolution*. Oxford: Oxford University Press.

Somanathan, E. 1991. "Deforestation, Property Rights and Incentives in Central Himalaya." *Economic and Political Weekly*, Vol. XXVI: 37-46.

The Social and Mental Dynamics of Cooperation

SOCOP

NSF Intellectual Merit and Social Impact Statement

Intellectual Merit

In its 2005 assessment of the top intellectual issues in contemporary science, the eminent journal **Science** included understanding the nature of human cooperation as among the leaders. There is currently no accepted model of how the cognitive capacities and motivational predispositions of humans interact to support cooperation and how these unique human traits evolved. While the social sciences have made major contributions to understanding cooperation, each behavioural discipline ignores a key part of the overall picture. The nature of cooperation is such that it does not break down into independent parts susceptible to isolated analysis by specific disciplines. The *SOCOP* project is a novel and uncompromisingly trans-disciplinary strategy for understanding human cooperation. We anticipate a number of specific scientific breakthroughs that promise to have practical application to problems of organizational governance, conflict resolution, and the promotion of intergroup tolerance.

Social Impact

SOCOP aims to identify both the pro-social and anti-social aspects of human cooperative tendencies, develop a new approach to organizational governance, public policy and economic incentives taking into account the cooperative nature of humans, and publish results in high-impact journals and international meetings. Until recently it was widely believed that promoting economic growth and social welfare was a matter of simply putting into place the appropriate economic institutions and making available the appropriate levels of investment resources. The human characteristics deemed important were cognitive and skill related alone. We now know that a high level of economic efficiency depends on such character virtues as honesty and commitment to hard work, and normative values including opposition to nepotism and corruption. These human characteristics are part of the general model of cooperation upon which SOCOP will elaborate in the course of our research

IP3: In Silico Evolution of Cooperation and Communication

Eörs Szathmáry

Institute for Advanced Study, Budapest

B.1. Individual Project contribution to the CRP

a. IP aims and objectives

Genetic evolution has endowed humans with the capacity of teaching, detailed imitation, a complex theory of mind, strong altruism and language. Genes must have influenced the nervous system in such a way that all these evolutionary changes became possible. There are billions of neurons, but only about 35 thousand genes. It is far from obvious how neuronal networks can be affected by relatively few genes in a highly indirect manner (through development) so that even complex cognitive traits show significant heritability. We take an *in silico* approach to the problem with aid of our Evolutionary Neurogenetic Algorithm (ENGA) that allows the selection of neuronal agents embedded in a simple or complex simulated physical and social environment. We seek to identify selective scenarios according to which complex communication or strong cooperation emerges, and we would like to analyze the correlated neuronal and genetic changes in the successful agents. Synergies between levels forms of cooperation and forms of communications will be sought for. The research will be carried out in close collaboration with Michel Kerszberg in order to better guide the evolutionary process and to successfully reverse-engineer the emerging networks. Since we are convinced that complex communication and complex cooperation go hand in hand, cooperative, simulated situations will be cast in a game-theoretical framework in order to see clearly and help cross-disciplinary and interactions. Since the communicating neuronal agents will not only learn but their learning capacity will undergo genetic evolution, our project is an eminent case of simulated gene-culture coevolution. Both these considerations strongly relate to the conceptual edifice of Herb Gintis.

Background: Young (and old) humans live in a completely artificial environment created by our predecessors, immediate and less immediate. The stimuli we are bombarded with are therefore highly special, i.e. they are strongly biased, partly by the social structure and partly by the productions of our parents. These biases, hence, are partly of genetic and partly of populational origin; they interact in complex ways with the direct biases that operate within the nervous system to give rise to cognitive development. It is on these fundamental theoretical considerations that our computer simulations must rest. The origin of language was identified as the last major transition (Maynard Smith & Szathmáry, 1995) with a genetic background. Language is a novel inheritance system with indefinitely large hereditary potential in humans. The question is how a transition from protolanguage (with only some word-like signals: Bickerton, 1990) to language is possible in a society of evolving agents. Transition

theory also calls attention to the effects of the division of labour and the emergence of higher levels of selection. It remains to be elucidated how all these conditions could have contributed to the human condition.

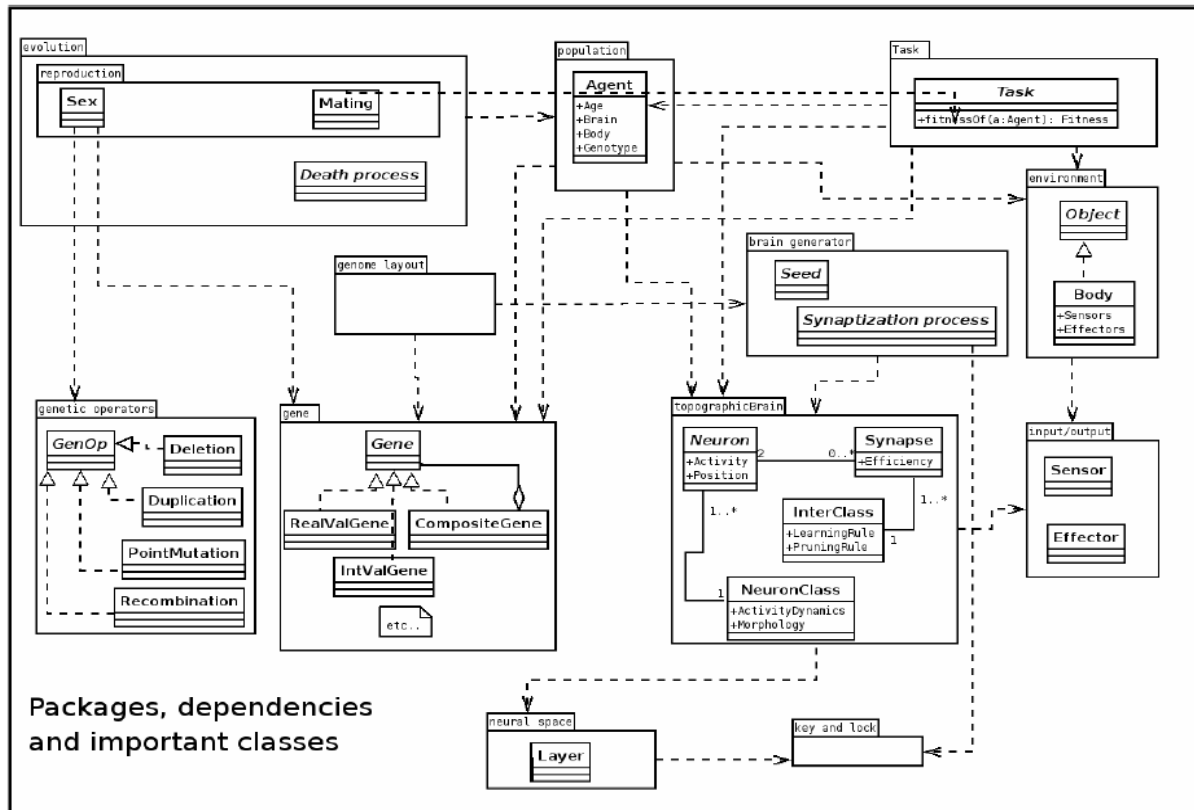
Some major transitions in evolution (such as the origin of multicellular organisms or that of social animals) occurred a number of times, whereas others (the origin of the genetic code, or language) seem to have been unique events. If all the extant and fossil species, which possess traits due to a particular transition, share a last common ancestor after that transition, then the transition is said to be unique. It is quite possible that there have been independent “trials”, as it were, but we do not have comparative or fossil evidence for them. What factors can lead to “true” uniqueness of a transition? A) The transition is variation-limited. This means that the set of requisite genetic alterations has a very low probability. “Constraints” operate here in a broad sense. B) The transition is selection-limited. This means that there is something special in the selective environment that can favour the fixation of otherwise not really rare variants. Abiotic and biotic factors can both contribute to this limitation.

It is hard to assess why language is unique. Even the “not enough time” case could apply, which would be amusing. But pre-emption, due to the subsequent cultural evolution that language has triggered, may render further trials very difficult indeed (Számádó & Szathmáry, 2006).

Previous Research:

Evolutionary neurogenetic algorithm (ENGA)

Our goal was the development of a new neuro-genetic system based on genetic algorithms and neuronal networks that allows coevolution of motor and communication behaviour and the neural architecture of the agents. We have been building an object-oriented software framework in order to



provide the required flexibility. Because of the interface system it is possible to change different modules of the software without complete redesign of the system. The framework has four main modules: (1) genome and genetic operators; (2) neuronal network representation; (3) genome layout (translator of the genome); (4) population and sexual reproduction. The central object is the agent interface. The overall structure of the framework can be seen on the Figure (Szathmary et al. 2006).

ENGA offers researchers a fine control over biological detail in their simulations. Our original intent was to create software with much potential for variability. That is, we wanted a piece of software which is general enough to allow for a wide range of experimentation but appears as a coherent system and does not fall apart into a loosest of unrelated pieces of code. This required careful specification and design; especially in partitioning it into modules and the specification of interfaces in a programme that has grown to about 90,000 lines of C++ code. In such a short communication it is impossible to acknowledge all researchers of all important input fields to this paper. We have been especially influenced by evolutionary robotics (Nolfi & Floreano, 2002), and by the evolutionary approach to neuronal networks with indirect encoding by Rolls and Stringer (2000). Our model is a

recombinant of these approaches, with some key new elements, such as topographical network architecture.

The software is organized into packages that are built upon each other, i.e. there is a dependency hierarchy between them. This gives the architecture a layered nature so that lower modules do not know about the existence of higher modules. The most important packages and their dependencies are shown in the Fig. Layered design allows easy modifiability of higher levels without the need to modify lower levels. Moreover, each layer exposes an interface that can be used by any client, even those deviating from the original purpose of simulating evolution of embodied communicating agents. The genetic module for example can be used in any evolutionary computation, not only those evolving artificial neural networks. We may as well talk about a multilevel software framework consisting of several modules that can be used individually or in combination with others to produce various kinds of evolutionary and neural computation related simulations. In the following sections individual packages are described in more detail.

c. Work plan

Selective scenarios

We propose that it is instructive to look at the various scenarios for the evolutionary origin of language, put them to a test of stringent criteria, and later use the result in an educated guess as to what preconditions one should put into models that could allow for the spontaneous emergence of language-like communication systems, including the use of symbols, compositionality and, ultimately, recursive syntax.

Selective scenarios differ in the way they neglect or reflect social organisation. In our experience the time during which one can carry out successful study with simulated neuronal agents is unpredictable. Thus we should aim a few studies we hope that we can really carry out given the limited time and resources: anything extra should be regarded as a bonus.

1. **Artificial selection for recursive abilities (syntax only).** Many regard recursion as the main or perhaps even only critical element human syntax. It is true that, for example, tamarin monkeys cannot master phrase structure grammar PSG (in contrast to finite state grammar FSG: Hauser & Fitch, 2004), but it is remarkable that apparently European starlings are also capable of pattern recognition in auditory input that confirms to central embedding (PSG: Gentner et al. 2006). Nobody knows how the necessary operations are carried out by nervous systems. In order to gain a partial answer, we shall expose the agents to selection conditions under which agents that can master PSG or FSG or both, will

be rewarded by high fitness. We shall analyze the correlated changes in the genome and the neuronal network of the simulated agents. The experiment can be carried out in a more or less passive manner. The first aims at comprehension (pattern recognition only). The second is more complex because it aims at production: in the completion task agents are rewarded when they are able to complete an output that follows PSG, for example. As language is a combination of comprehension and production, both sides should be analyzed.

2. **Natural selection for cooperation with communication (syntax with semantics).** It seems that an activity such as cooperative hunting (Számadó & Szathmáry, 2006) could have likely been the context in which complex communication was selected for. In this experiment agents gain a fitness advantage when they successfully hunt for different kinds of simulated prey, and communication confers advantages in the game. Naming of prey should emerge because it will be detrimental not to “know” what kind of dangerous prey they are hunting for, and not everybody will be able to see directly what is going on all the time. Furthermore, certain aspects of the situation will be communicable only using PSG rather than FSG. Design of such a scenario is an art in itself, and we are likely to be forced to break it down into more elementary situations that the agents can solve. We see this experiment as a culmination of our research strategy.

d. Deliverables/milestones

The product of this research will be a deeper understanding of the possible genetic control and neuronal realisation of specific human capacities to cooperate and communicate. The main results will be published as research papers in high-ranked journals, but a more synthetic summary, possibly in book form, is also foreseen.

e. Works cited

- Bickerton, D. (1990) *Language and Species*. University of Chicago Press.
- Fitch, W.T., Hauser, M.D. (2004) Computational constraints on syntactic processing in nonhuman primates. *Science* 303, 377-380.
- Gentner, T.Q. et al. (2006) Recursive syntactic learning by songbirds. *Nature* 440, 1204–1207.
- Hauser, M.D., Chomsky, N. Fitch, W. T. (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569-1579.
- Maynard Smith, J. & Szathmáry, E. (1995) *The Major Transitions in Evolution*. Oxford University Press.
- Nolfi, S., Floreano, D.: Synthesis of autonomous robots through evolution. *Trends Cogn. Sci.* **6** (2002) 31-37.

Rolls, E.T., Stringer, S.M. (2000) On the design of neural networks in the brain by genetic evolution. *Prog Neurobiol* **61**, 557-579.

Számadó, S. & Szathmáry, E. (2006) Competing selective scenarios for the emergence of natural language. *Trends Ecol. Evol.* in press.

Szathmáry, E., Szatmáry, Z., Ittész, P., Gergő, O., Zachár, I., Huszár, F., Fedor, A., Varga, M. & Számadó, S. (2006) *In silico* evolutionary developmental neurobiology and the origin of natural language. In: (Lyon, C., Nehaniv, C. L. & Cangelosi, A., eds) *Emergence of Communication and Language*. Springer, Artington. In press.

IP 4: Proximate explanations for human and animal cooperation

Cristiano Castelfranchi

Istituto di Scienze e Tecnologie della Cognizione, ISTC-CNR

B.1. Individual Project contribution to the CRP

a. IP aims and objectives

How do human decisions concerning cooperation differ from those of other primates? What are the cognitive and emotional prerequisites for widespread cooperation in humans? How did human cooperative behaviour evolve? To tackle these questions, the main focus of this IP will be on *the theoretical and experimental understanding of the proximate causes of cooperative behaviour both in non-human and human primates* in situations involving many agents. The existing literature does not include detailed models of the psychological mechanisms behind cooperation, and in the animal literature, data on multi-player settings have not been collected. We will do this.

This operational (proximate explanation) analysis compliments the experimental (behavioural data) and evolutionary (ultimate explanation) of other Principal Investigators. Several of these mechanisms are predicted or presupposed by the research on the institutional niche construction and will be investigated in collaboration with Bowles. The empirical investigation on the development of cooperative norms undertaken by Mace will also provide a test bed for our model of norm-compliance. We will also investigate how the proximate neural mechanisms advanced by Kerszberg could enable the cognitive mechanisms explored in this IP. Finally, in collaboration with Gintis, we will deploy the BPC framework to strengthen the interdisciplinary compatibility of our models and results.

We are an interdisciplinary group of cognitive and social scientists from cognitive psychology and distributed artificial intelligence (Prof. Cristiano Castelfranchi, Dr. Rino Falcone and Dr. Maria Miceli), social psychology and agent based simulation (Dr. Rosaria Conte and Dr. Mario Paolucci), cognitive primatology (Dr. Elisabetta Visalberghi, Dr. Francesco Natale, Dr. Elsa Adessi and Dr. Gabriele Schino), and evolutionary economics (Prof. Ugo Pagano).

Reciprocity in non-human primates: partner selection in Cebus apella and other primates

It has been suggested that nonhuman primate reciprocal altruism may be based more on the triggering of “positive” emotions by the receipt of altruistic behaviours than on any mechanism of book-keeping (de Waal 2000). This hypothesis has rarely been explicitly tested (exceptions see Silk et al., 2005; Melis et al., 2006; Jensen et al., 2006), and then only in dyadic contexts. We will test this in larger groups in more natural settings in which partner choice plays a significant role. Experiments in this

area will be complemented by ethological captivity and field studies. (see Schino, in press, for a recent meta-analysis).

Reciprocity in humans: reputations, gossip, and norms

Two crucial mechanisms are available only to humans that might be responsible for the evolution of our astonishing level of cooperation: a) reputation and b) norms.

Reputation and reciprocal altruism

Reputation is a type of social knowledge that adds cohesiveness to social groups and allows for distributed social control and sanction (Barkow, 1992; Greif, 1989). It is a property that unwilling and unaware individuals derive from the transmission and manipulation of social evaluations, and that contributes to regulate natural societies by promoting prosocial behaviours (Alexander, 1987; Conte and Paolucci, 2002; Dunbar, 1996). Reputation and its transmission, i.e. gossip, allows people to use indirect informational reciprocity in partner selection and social control in large open non-kin settlements, where agents can be preceded by their reputation. Despite its role in human societies, how this critical type of knowledge is processed in the agents' minds, how social structures and infrastructures are related to it, and consequently how it affects agents' behaviour has not yet been clarified. This IP will clarify these mechanism. Why do people exchange information? Under what conditions is gossip about a conspecific is similar to gossiping about an event? What are the costs and the benefits of these two kinds of informational exchange? We will address these questions.

Norms and Strong Reciprocity. Explaining how cooperation in anonymous, one-shot encounters may have evolved is still an unresolved puzzle. One explanation is gene-culture co-evolutionary process (Gintis et al. 2004), Henrich et al., 2003 and Bowles in this project). This ultimate evolutionary explanation rests on the assumption that agents can be motivated to act as strong reciprocators (*conditional cooperators* and *altruistic punishers*) by compliance with a norm (Bowles and Gintis, 2004; Fehr and Fischbacher 2004).

This IP will explore the “psychology of norm compliance” by unravelling the cognitive pre-requisites to comply with a norm. We will explore how ‘respect for the authority’ can function as a cognitive process leading to the adoption of particular course of action and the special way in which norms are processed in the practical reasoning of the agents (Conte and Castelfranchi, 1995). Secondly, two aspects of punishment will be studied. It is contended that strong reciprocators are altruistic punishers (Bowles and Gintis, 2002). The current most accredited theory of the mechanism behind norm violators detection postulates an innate cheater detector module evolved for this task (Cosmides 1989; Tooby and Cosmides, 1992). We will advance an alternative and more parsimonious hypothesis showing that violation detection can be a by-product of the human capacity to reason teleologically

(means-end reasoning). A set of experiments with human subjects will be designed to test this hypothesis. As far as the cost of punishment is concerned, a very cheap and effective punishing strategy is provided by “blaming” somebody for having violated a norm (Strawson 1962). Blaming is considered as a form of moral aggression and a very effective one to enforce moral norms (like the norm of conditional cooperation underlying strong reciprocity). The peculiar way blaming functions to enforce conformity will be analysed. Two strategies of blaming (direct and indirect forms of blaming) will be modelled and their effects disentangled.

The evolution of human cognitive mechanisms for cooperation

We will explore several possible hypotheses about the evolutionary origins of the human cooperative mind with its sophisticated cognitive abilities. In particular we will investigate (a) the possible co-evolution of authority and subject psychologies underlying norm issuing and norm compliance (b) the role of deontic teaching in cultural transmission and (c) the peculiar role of the human fertilization system in the evolution of human mechanisms for complex cooperation.

As for the first sub-task, we will explore two possible co-evolutionary processes leading to the emergence of authority and subjects psychologies. One plausible hypothesis is that such psychologies rest on more ancient submission/dominance adaptations (Cummins, 1998). The second one sees norm compliance as a kind of docility similar to the one we find in learning from teaching (Simon, 1990). The plausibility of these two evolutionary hypotheses will be theoretically and formally explored.

As for the second sub-task, we will explore “deontic” teaching as a different process in cultural transmission. While transmission is usually modelled as a passive process (the agent imitates or copies what the others do), there is also a different process of active transmission (the imitated agent aims at showing how to behave to the imitating agent). The motivation for active teaching can be seen also as a consequence of being an altruistic punisher.

Finally, we will explore the hypothesis that human sexual selection explains the evolution of those mechanisms that are to a large extent human-specific, such as political skills, the relatively successful management of common property, and prosocial emotions. The unique human fertilization system, based on female concealed ovulation and female selective mating choices, sets a system of rewards in the domain of sexual selection that favoured the formation of a type of quasi-monogamic relations based on long term commitments and trust. In turn, the evolutionary importance of these relations select in favour of individuals endowed with high emotional and intellectual faculties.

b. Methodologies/Experiments

Theoretical and computational modelling: For objectives (2) and (3), detailed computational models of the mechanisms will be developed as a way to defend the soundness of the proposals. As for objective (1), experiments will explore reciprocation and its time-frame in captive capuchin monkeys (*Cebus apella*). As for objective (2), the experiments with human subjects aim to falsify the current dominant hypothesis as far as norm violation detection is concerned. A set of laboratory experiments on the model of the Wason Selection Task will be designed to show that the systematic errors in reasoning with conditionals are resolved when subjects are capable of framing the problem from a teleological perspective. The study of the evolutionary dynamics producing specific proximate causes (objective 3) will be approached by means of mathematical modelling and agent-based simulation (ABS).

c. Work plan

Year 1: During the first year, the experiments with the non-human primates and with humans will be initiated. All the preliminary results on the three objectives (theoretical, mathematical and computational models) will be published in the form of conference papers and articles on high-level peer-reviewed journals.

Year 2: During the second year, in addition to the developments in the scientific results and dissemination (conference papers and journal articles), an interdisciplinary conference on the topics of this IP will be organized to favour the integration and the collaboration within and outside the CRP.

Year 3: In the last year, the theoretical, computational, mathematical and experimental scientific results of the IP will be collected in an edited book to be submitted to a major publishing house at the international level.

d. Deliverables/milestones

During the first two years of the project, the main deliverables of this IP will be in form of conference papers and articles on peer-reviewed journals. At the end of the year all the project contributions will be collected in a scientific report (*D1* at the end of Year 1 and *D2* at end of Year 2). During the second year, we will also organize an Interdisciplinary Workshop on “Understanding the proximate causes of cooperation” which will count both as a deliverable (*D3*) and as the main project milestone. Finally, at the end of Year 3, the edited book will be the final deliverable (*D4*) of this IP.

e. Justification for budget items, especially staff costs and equipment

2 Post-docs and 1 Student Assistant will be hired to work in this IP. The post docs are needed to conduct high-level and independent research resulting in journal articles and book chapters. The first Post-doc will be in charge of managing a complex research project on “Proximate causes for Strong Reciprocity” while the second Post-doc will be in charge of a research project on “Proximate causes for Reciprocal Altruism” with a special emphasis on reputation and gossip. These researchers should already be able to conduct independent research both at a theoretical and an experimental level. They will be also in charge of coordinating these activities with other IPs to explore the compatibility of the investigated proximate explanations with the ultimate evolutionary explanations. Given the interdisciplinary aims of this CRP, these researchers will be also capable of strengthening the interactions with other IPs and should be already able to acquire and transfer technical knowledge from one domain of expertise to others.

The Student Assistant will support the execution of the experiments with animals and will be trained to conduct research on these issues. Building new research expertise in this area is particularly important to influence future interdisciplinary research.

- Travel costs

Particular weight in this voice is necessary to support observational fieldwork to be done in Brazil for the whole project duration. The remaining part of the funding will be used to attend to conferences, project internal meetings and CRP interdisciplinary workshops and to support short-term visits and exchanges between CRP partners.

- Equipment Regular costs to update the technological equipments of the staff.

f. Works cited

- Alexander, R.D. (1987). *The Biology of Moral Systems*. Aldine de Gruyter, New York.
- Barkow, J.H. (1992). Beneath new culture is old psychology: Gossip and social stratification. In J. Barkow, L. C. and Tooby, J., editors, *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, pages 19–136. Oxford University Press.
- Bowles, S. and Gintis, H. (2002). Homo reciprocans: Altruistic punishment of free riders. *Nature*, 415(10):125–128.
- Bowles, S. and Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, 65:17–28.
- Boyd, R. and Richerson, P. (1985). *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.
- Conte, R. and Castelfranchi, C. (1995). *Cognitive and Social Action*. UCL Press, London.

- Conte, R. and Paolucci, M. (2002). *Reputation in Artificial Societies*. Kluwer, Boston.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? studies with the wason selection task. *Cognition*, 31(3):187–276.
- Cummins, D.D. (1998). Social norms and other minds. In D.D. Cummins and C. Allen (eds.), *The evolution of mind* (pp. 30-50). New York: Oxford University Press.
- Dunbar, R. (1996). *Grooming, Gossip and the Evolution of Language*. Faber and Faber Limited, London
- Fehr, E. and Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, 8(4):185–190.
- Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90:980–994.
- Gintis, H, Bowles, S., Boyd, R. and Fehr, E. (2005), *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life*. The MIT Press, Cambridge, MA 2005.
- Greif, A. (1989). Reputation and coalitions in medieval trade: Evidence on the Maghribi traders. *The Journal of Economic History*, 49 (4), 857-882.
- Henrich, J., and Boyd, R. (2001). Why People Punish Defectors: Weak Conformist Transmission Can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas. *Journal of Theoretical Biology*, 208, 79-89.
- Henrich, J., Bowles, S., Smith, E. A., Young, P., Boyd, R., and Sigmund, K. (2003). The cultural and genetic evolution of human cooperation. In Hammerstein, P., editor, *Genetic and cultural evolution of cooperation*. MIT Press.
- Jensen K, Hare, B, Call, J., and Tomasello, M. (2006). What's in it for me? Self-regard precludes altruism and spite in chimpanzees. *Proceedings of the Royal Society of London*, B273(1589), 1013-1021.
- Kurzban, R. and Houser, D. (2005) An experimental investigation of cooperative types in human groups: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences*, 102(5), 1803-1807.
- Melis, A.P., Hare, B. and Tomasello, M. (2006) Chimpanzees recruit the best collaborators. *Science*, 311, 1297-1300.
- Schino, G. (in press) Grooming and agonistic support: a meta-analysis of primate reciprocal altruism. *Behavioural Ecology*.
- Silk, J.B., Brosnan, S.F., Vonk, J., Henrich, J., Povinelli, D.J., Richardson, A.S., Lambeth, S.P., Mascaró, J. and Schapiro, S.J. (2005) Chimpanzees are indifferent to the welfare of unrelated group members. *Nature*, 437, 1357-1359.
- Simon, H.A. (1990). A mechanism for social selection and successful altruism. *Science*, 250, 1665-1668.

Strawson, P.F. (1962) Freedom and Resentment. In "Freedom and Resentment", London: Methuen, 1974.

Tooby, J. and Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. C. and Tooby, J., editors, *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, pages 19–136. Oxford University Press.

de Waal, F.B.M. (2000) Attitudinal reciprocity in food sharing among brown capuchin monkeys. *Animal Behaviour*, 60, 253-261.

B.3. Related Projects

- MINDRACES “From Reactive to Anticipatory Cognitive Embodied Systems” (Strep Project Contract N. IST-511931; VI FP “Cognitive Systems”) for the role of anticipation in mind and that of expectations in the evolution of conventions and norms.
- CONTACT “Consciousness in interaction” (ESF CRP under the CNCC call) for the social origins of consciousness and the interplay between cooperation and self-consciousness.
- HUMAINE (Network of Excellence; IST-2002-2.3.1.6 Multimodal Interfaces) on modelling and understanding emotions.
- SEDSU “Stages in the Evolution and Development of Sign Use” (VI FP NEST “What it means to be humans”) for the understanding of the cognitive capacities for symbolic understanding.
- ANALOGY “Humans – The analogy-making species” (VI FP NEST “What it means to be humans”).
- E-REP “Social Knowledge for e-Governance” (Contract N. KPS-28575; VI FP “Citizens”) for the design of reputation technology.

B.4. Relevant international scientific relationships

University of Vienna, University of Edinburgh, University of Kyoto, Max-Planck Institute at Leipzig, University of Georgia, Athens, University of S. Paulo, University of Surrey, Centre for Research on Social Simulation, Universität Koblenz-Landau, Manchester Metropolitan University, Centre for Policy Modelling, Goldsmiths College, CNRS-Marseille, University of Portsmouth, Lund University, University of Leicester, University of Tel Aviv, University of Louisiana, Fundacao BioBrasil.

B.5. Extra item for additional information required by a national funding agency

CNR has confirmed that the present Individual Project satisfies the eligibility requirements.

The outline proposal incorrectly specified Prof. Pagano from University of Siena as a Principal Investigator. This mistake has been amended and he now correctly figures as a sub-contractor of the present IP. Prof. Pagano expertise in mathematical modelling and evolutionary economics is in fact a necessary requirement to meet the third objective of the present project and hence he will be responsible for this sub-task as specified above.

IP5: High-Level Neuronal Architectures for Cooperation and Communication***Michel Kerszberg*****Université Pierre et Marie Curie****B.1. Individual Project contribution to the CRP****a. IP aims and objectives**

We aim to discover, by the combination of rational and evolutionary design of neuronal networks, what kinds of neural network properties allow the emergence of cognitive/computational capacities that underlie various forms of cooperation and communication among simulated agents. For cooperation, we would like to demonstrate that under appropriate selective conditions, and with appropriate neural and populational preadaptations such as planning and decisional abilities, the emerging architectures are capable of recognizing individuals (from behaviour or other external features) and that the agents become genetically biased towards strong altruism. The architectures emerging from this coevolution of genes and culture (Gintis IP) are the proximate causes (Castelfranchi IP) for the altruistic behaviour, while the necessary reciprocity implies the existence of appropriate niches where the behaviour becomes a norm (Bowles IP). We aim to show that a highly coordinated, synergistic collaborative task such as cooperative hunting, leads to agents with shared attention/intentionality and that without emerging communication the task cannot be fulfilled as efficiently. The semantic content of the communication, constrained by the cooperative situation, is expected to lead to not only symbolic reference but also to at least rudimentary syntax. Reverse engineering of the emerging architectures can lead to testable propositions about the human condition.

b. Methodologies/experiments

Background and Previous Research: The way in which biological, and in particular neural architectures are related to information processing abilities is subtle. It has been shown (Kerszberg 2004) that, contrary to what is sometimes glibly asserted, the coding capacity of the genetic system is huge, leaving open the question of why certain abilities are «innate» while others «need» to be «learned», sometimes with input from congeners, pedagogy, work sharing and cooperation. In spite of much work, e.g. by Edelman's group, little progress has been evident. In part this is due to the excessive cost of building actual robots to perform the work, when the necessary environment can much more cheaply be simulated.

We have introduced (Kerszberg 1989) and developed previously (Kerszberg, Dehaene and Changeux 2001) a general framework within which the interactions between intrinsic brain developmental factors

and environmental stimulatory factors (including congener behaviour) can be studied. This framework has been applied recently to the study of human performance in the so-called Stroop task (involving the forcing of an «effortful» response, DKC), and delayed-response planning and switching in delayed-response task performance by non-human primates (Gisiger, Kerszberg and Changeux, 2004 ; Gisiger and Kerszberg, 2005).

c. Work plan

Modelling Research Program: The characters inherited from our parents shape our capacity to survive, reproduce and transmit those characters. The characters of present living beings are thus the result a long process of selection. While selection decides among possibilities, however, it is up to the laws of organism (here, brain structure) development to propose what the available possibilities actually are. In the central nervous system (CNS), how behavioural variations are generated is an enormously difficult question, because the development of brain structure and wiring is highly complex, and the relation of wiring to behaviour is anything but straightforward. How did the brain of the higher primates come to offer the possibilities which, when selected, became our most exclusive and prized features? This is the question we want to address.

While many of the neural operations performed daily by humans are not exclusive to them as a species, some obviously are, although it is not simple to discern exactly which. Here we focus on two obvious candidates. I consider those from the point of view of neurobiology.

Choice and planning: Leaving aside the philosophical issue as to whether we actually are free in our choices, we do choose between alternatives and these choices shape our personal lives. There is wide agreement that while all primates may display a capacity for choice, when put in the right experimental conditions, the extensive pondering of alternatives “in the wild” is a human attribute.

Using symbols: While higher primates may have been trained in the use of various “symbols”, the use of symbols in the full sense of the word is a human attribute. One of the oldest uses of symbols is in the economy, first with exchange, then with money. The major innovation which at once uses symbols and is made possible by them is of course language. It has been argued recently (Chomsky, Hauser) that the key point in language is recursion, *i.e.* the possibility for symbols to symbolise symbols (including themselves), *ad infinitum*. A related point is multiple instantiation of symbols (In the recursion “I see that you see him”, the verb “to see” appears twice).

Evolving choice, symbols and recursion: proposed ways towards solutions: It would not be honest to claim that we are going to solve these age-old questions in a few years. The best we can do is offer and

explore some pathways towards progress, some quite novel, others less so but deserving to be revisited. What are the genetic-developmental features of the brain that might be brought to bear in the generation of selectable behavioural variations? Here our hope is the theoretical works of Kerszberg and Szathmary IP's will become able to guide in part and predict some of the outcomes of the genetic and genomic work described in Navarro's and Fehr's IP's.

The neural development mechanisms we want to invoke

Working memory systems

There are well established models implementing working memory (WM). The one we use is from Kerszberg, Dehaene and Changeux 1992 (KDC), based on Hebbian plasticity. In this context however, we want to insist on the necessity of reset mechanisms in working memory. Once you have finished a job, you can forget what you had in mind. This is usually glossed over, but is actually *capital* for proper operation of WM within a network which is supposed to perform a planned task. The correct reset at the correct time is something that must evolve and/or be learned; at any rate it is part of the neural requisites for the task.

Hierarchies of memory systems

Genes “easily slip” into building repetitive structures (such as cortex); thus we assume WM too can be demultiplied, and become organized hierarchically, which is fundamental to ENGA (see Szathmary IP). While WM à la KDC, in addition to remembering inputs for a time, can detect simple logical (space-time) correlations between some inputs, with a hierarchical organization we can quickly build a system where there is a neural correlate of just about any stimulus or situation (grandmother neurons).

Gating systems

In a washing machine, the drum should not start rotating before there is water in it. In the CNS there is the same necessity of coordination. In Gisiger *et al* (2005) we have proposed the existence of gatings that allow information to circulate in various places and directions, and of resetting systems (see above) which make parts of the system ready for the job next at hand. These various “executive” systems play a major role in organising complex, planned behaviour from elementary component behaviours. Just like resets, gatings are part of the functioning of the system, and should not be taken for granted.

Inner reward systems

In further structuring behaviour, inner rewards plays a major role and neural correlates for these have been found indeed. In essence, we are talking here about the satisfaction we feel when we have done something “right”. While this may be connected to external reward (particularly in infancy) it is certainly different from external reward. The unease or surprise we feel when we hear a nongrammatical sentence is a negative internal reward; in a child, the fact that others seem to understand him (i.e. they give him what he wants) is a positive reward. This is difficult to define precisely, and is connected with aesthetic sense for instance; yet it appears to be a fundamental element in checking individual progress.

Recursion system

Implementing the recursion system is the most challenging problem of all. Here we think that a more careful analysis of how related pieces of information are bound together in the CNS might be fruitful. In particular, the way this is usually assumed to take place is through a time-synchronisation of the relevant neural systems; and certainly, this allows gaining quite a bit of flexibility in connecting various unrelated mental objects such as the *signifiant* and *signifié* (to adopt Ferdinand de Saussure's linguistic nomenclature); but obviously, more complex and subtle phenomena must take place when a given symbol is reused without its various instantiations becoming hopelessly mixed together. We plan to test various neurobiologically plausible mechanisms which may contribute to progress on this extremely difficult issue.

How these mechanisms may contribute to a solution

Brain development “by itself”

The mechanism proposed in (Kerszberg 1998) will permit us to build a model that allows a factoring of the genetic and epigenetic effects in neural development. In this model the effect of genetic components is subsumed by the existence of differential identification and probabilistically biased interconnection between neural units (neurons, areas, etc.); development then introduces interactions between these units, first in the form of spontaneous activity, and later, stimulus-evoked activity. Because of the genetic biases of the interconnections, spontaneous activity is itself biased and the resulting plastic changes of the connections, which can be called maturation, will also be biased. Evoked activity will be responsible for further plastic modifications, which may come under the heading of learning. Of greatest impact for human cognitive development however are a different set of factors, familial and populational.

Brain development in family and population

Young humans (and old) live in a completely artificial environment created by our predecessors, immediate and less immediate. The stimuli we are bombarded with are therefore highly special, i.e. they are strongly biased, partly by the social structure partly by the productions of our parents. These biases, hence, are partly of genetic and partly of populational origin; they interact in complex ways with the direct biases discussed in the previous paragraph to give rise to cognitive development. It is on these fundamental theoretical considerations that our computer simulations must rest.

d. Deliverables/milestones

The product of this work will be a deeper understanding of the neural structures underlying the human capacities for cooperation and communication, and their possible genetic control and evolution. It is hoped that our results will have predictive value for the experimental behavioural studies by Fehr and Navarro, and will interact with the theoretical approaches of Bowles and Castelfranchi, while helping to resolve some difficult issues in the unification of the behavioural sciences (Gintis). The main results will be published as research papers in high-ranked journals, but a more synthetic summary of the results of this and Szathmary's IPs, possibly in book form, is also planned.

e. Justification for budget items

The project is computationally very intensive, programming will require highly skilled programmers, and the many parallel computer experiments need experts to run and evaluate them. Efficient interfacing with other workpackages also needs time and effort. Computer equipment required (in particular for compatibility with Budapest) in addition to the workstations already available locally, namely a scalable Linux processor cluster, will cost about 80.000 €. Our research unit is asking CNRS, independently, to fund the hiring of an «ingénieur de recherche» to deal with the constraints of running the equipment. We should need 20.000 € for travel and meetings within the collaboration, bringing the total to 100.000 €.

f. Works cited

Kerszberg, M (2003) Genes, neurons and codes: remarks on biological communication. *Bioessays* 25(7):699-708.

Kerszberg, M (1990) Genetics and Epigenetics of Neural Function: a Model. *J. Cog. Neurosci.* 2:51-57.

Kerszberg, M, Dehaene, S and Changeux, JP (1992) Stabilization of Complex Input-Output Functions in Neural Clusters Formed by Synapse Selection, *Neural Networks* 5:403-413.

Dehaene, S, Kerszberg, M and Changeux, JP (1998) A neuronal model of a global workspace in effortful cognitive tasks. *Proc Natl Acad Sci U S A*. 95(24):14529-34.

Gisiger, T, Kerszberg, M and Changeux, JP (2005) Acquisition and performance of delayed-response tasks: a neural network model. *Cereb Cortex*. 15(5):489-506. Epub 2004 Sep 1.

Gisiger, T and Kerszberg, M (2006) A model for integrating elementary neural functions into delayed-response behaviour. *PLoS Comput Biol* 2(4):e25. Epub 2006 Apr 7

IP6: Exploring the behavioral genetics of Trade and Cooperation

Arcadi Navarro

Universitat Pompeu Fabra

B.1. Individual Project contribution to the CRP

a. IP aims and objectives

a.1. Overview

Modern humans are the only primate species to show extraordinary variation in social organisation, kinship, political and trade systems, expression and application of symbolism and so on. These wide range of behaviours includes two crucial human traits: intricate cooperative behaviours and the ability to develop complex trade systems. Such traits have been key to the ability of humans to adapt to many different environments and changing circumstances along our evolutionary history. What is the degree of phenotypic diversity in these traits? Do genetic factors contribute to such diversity in these traits? If so, how many factors? What may their relative contributions be? How may they be interacting among them and with environmental contributions? Which may be their genetic architecture and how do they differentiate humans from other primates?

We aim to conduct research on this topics in close collaboration with other IPs in this CRP. In particular, we intend to develop and select behavioural variables in coordination with the groups of Sam Bowles (Santa Fe, USA), Ruth Mace (UCL, UK), Cristiano Castelfranchi (CNR, Italy) and Ernst Fehr (University of Zurich, Switzerland). Behavioural experiments, sample collection and genetic analysis will be performed in close collaboration with Ernst Fehr. In addition, results of our research will provide insight into the genetic architecture of behavioural traits that will contribute to the production of more precise models by the aforementioned groups and by those of Eörs Szathmáry (Eötvös Loránd University, Hungary) and Michel Kerszberg (UPMC, France).

a.2. State of the Art

Nowadays, variability in complex behavioural traits can be measured with detail. First, exhaustive questionnaires are available to register self-reported behavioural measures of traits such as altruism, empathy, nurturance or aggression [ref 9, 10]. Second, these traits can also be measured by means of rating by peers or other individuals (Brendgen et al. 2005). Finally, laboratory experiments (social games such as the ultimatum game or the prisoner's dilemma) can provide direct measures of an individual's behaviour under controlled circumstances (Maynard Smith 1982). All these instruments

are providing unprecedented opportunities to obtain reliable, stable and consistent measures of human behaviour (Bouchard & Loehlin 2001).

There is an increasingly compelling amount of evidence that behavioural differences between humans could be, in part, the outcome of genetic variation, just as it is the case for many other complex traits. Such evidence comes from several different research lines. First, different levels of a number of hormones have been correlated with different behaviours. For example, a significant relationship exists between aggressive behaviour and testosterone level (Harris 1999) and levels of testosterone are known to be, in part, due to the genetic makeup of individuals (Jiang & Huhtaniemi 2004). Levels of serotonin have also been shown to correlate with different prosocial behaviours and perceptions (Rot et al. 2006; Zaboli et al. 2006). Of particular interest to the evolution of cooperation and trade is a recent study from the group of one of the associates to this project which, by means of a social game with monetary payoffs, showed that intranasal administration of oxytocin causes a substantial increase in trust among humans by affecting an individual's willingness to accept social risks arising through interpersonal interactions (Kosfeld et al. 2005). A second line of evidence for a significant role of genetic factors in human cooperative behaviour comes from the study of twins. Recent experiments with heterozygous and homozygous twins show a much larger correlation of behaviours among homozygous twins (Rushton 2004; Brendgen et al. 2005; Kosfeld et al. 2005). A third line of evidence comes from the study of mental disease. Genetic variants have been linked to many personality disorders, ranging from autism to suicidal behaviour or drug addiction (Van Belzen & Heutink 2006; Van den Bogaert et al. 2006; Zaboli et al. 2006). Finally, even if the current degree of genetic variability in behaviour is unknown, it does make perfect sense to assume that behavioural diversity of genetic origin must have been crucial in the evolution of human-specific cooperation abilities since, as shown by classical population genetics (Fisher, 1958), natural selection can only work upon genetic variance.

All this evidence coming from humans is reinforced with much more detailed knowledge about the genetic basis of animal behaviour. Perhaps the best known example is the genetic basis of the different mating strategies of the prairie and pine voles, regulated by different variants of the vasopressin receptor gene (Lim et al. 2004; Hammock & Young 2005). There are, however, many other examples, including cases in which the behaviour of an individual within a group depends on its genotype. This is the case of saddleback and moustached tamarins, in which genetically-based colour vision status affects spatial positioning in the group. Trichromats tend to get further from their neighbours than their dichromatic conspecifics, purportedly because they differ in their perceptions of predation risk (Smith et al. 2005).

Such detailed knowledge upon the genetic factors underlying standard behavioural variability within a species is not available for humans. This is paradoxical, not only because humans are the species for which behavioural studies are more abundant and detailed, but also because the study of the genetic architecture of complex traits has registered enormous advances over the last decade. The paradox becomes even more striking if one considers the great deal of resources is being devoted to the study of the genetic basis of many complex phenotypic traits (such as the concentration of clotting factors Sabater-Lleal et al. 2006), and many complex diseases, including traits and diseases for which the evidence of a role of genetic factors is weaker than for many behavioural traits (Petronis 2006; Toga and Thompson 2005 and references therein)

Given the general difficulty in obtaining large pedigrees, genetic studies increasingly rely in a population-based set of techniques under the common name of association analysis (Collins et al. 1999; Risch 2000; Cardon & Bell 2001; Marchini et al. 2005). These techniques allow the detection of genetic variants associated to any given trait by the simultaneous study of hundreds or thousands of individuals. Association analysis is based in the fact that, in a given population, the different alleles of close loci are organized in chromosomes in a non random way. That is, allelic variants of loci that are close-by in a chromosome tend to be correlated. Then, if are points of the genome containing allelic variants that are responsible for phenotypic effects and, simultaneously, there are points in the genome where it is easy and cheap to detect which alleles are present, then we can use the later ones (named *marker loci*) to estimate and/or locate the former ones (named *causal loci*).

The cheap and easy-to-use key that opens the door of association studies are SNPs (Single Nucleotide Polymorphisms). Millions of these variants affecting a single nucleotide have been described in the human genome (<http://www.ncbi.nlm.nih.gov/>; <http://www.hapmap.org>), they are quite evenly distributed and techniques to genotype them in an industrialized, cost effective way have been developed. In particular, microarray-based technologies allow to perform "Whole Genome Scans", genotyping studies covering the whole genome with up to hundred of thousands of marker SNPs. This is increasingly becoming the choice approach to quickly detect any genomic regions that may play a role in the phenotypic variability associated with any complex trait, since it allows not only for the detection of simple causal genetic variants (such as point mutations changing a single aminoacid in a protein), but also for the detection or more complex causal variants, such as Copy Number Variants (generating individuals with different copy numbers of given genes, see Sharp et al. 2006a and 2006b).

a.3. Objectives

The central goal of this *Individual Project* is to provide a first genome-wide scan of the genetic factors influencing cooperative and trade behaviour.

Questions to which this study aims to contribute are, among others, what is the current range and potential genetic basis of human variability in cooperative behaviour? How many loci and with what distribution of effects may be involved? What does it imply in the context of the evolution of cooperation and trade?

b. Methodologies/experiments

We intend to proceed in two steps:

i) Phenotypic study.

i.1) Develop a set of consistent measures of cooperative and trade-related traits to be taken in social games. As for any phenotypes, measures have to be easy to take, reliable, consistent and stable within individuals and across time and games. Ideally, measures should be taken as scores in games that are well-known in experimental psychology and experimental economics (such as the ultimatum game or the sequential prisoner's dilemma) and for which individual responses have been studied in detail. If necessary, preliminary social games will be run during this phase upon limited sets of individuals to ensure consistency and replicability of scores. The development of these measures will be carried out in collaboration with other IPs in this CRP, specially those that intend to model or measure cooperative and trait behaviors, such as the S. Bowles, R. Mace, C. Castelfranchi, E. Fehr and E. Szathmari,

i.2) Perform the choice laboratory experiments upon a sample of at least 600 individuals and score their results. The individuals' behaviour will also be assessed by means of a comprehensive battery of behavioural questionnaires. The comparison of games, questionnaires and individuals will allow the characterization of the choice behavioural variables to include in our study. Individuals will be chosen from a single population to avoid stratification problems. Experiments will be carried-out both at the University of Zurich (in collaboration with Ernst Fehr) and the Universitat Pompeu Fabra, if necessary with replicas at different sites to ensure consistency of results.

ii) Genetic study.

ii.1) Genotyping. DNA samples from the individuals taking part in the games will be obtained by means of standard mouth-swabs. DNA samples will be processed and genotyped using the facilities of Spain's National Genotyping Center (<http://www.cegen.org>). Affymetrix's GeneChip Mapping 500k arrays will be used.

ii.2) Data analysis. Data analysis will be performed by means of SNPator, a web-based environment oriented towards statistical genomics developed by us (<http://bioinformatica.cegen.upf.es>). Associations of SNPs, clusters of SNPs and potential CNVs with the measured behavioural traits will be studied. The usual statistical precautions to ensure correction for multiple tests will be taken.

Haplotype-based studies of the potential recent influence of natural selection upon candidate genomic regions will be performed.

ii.3) Functional analysis. We will perform functional analysis of the gene content of genomic regions showing stronger association with the measured behaviours. This analysis will be performed by means of public databases linking genome annotations with functional information, such as Gene Ontologies (<http://www.geneontology.org>). Within these regions, we will try to locate genes, regulatory elements or RNAs that may be candidates for further dissection in order to describe the exact variants contributing to human behavioural variability.

ii.4) Comparative analysis. In parallel to the functional analysis above. We will study the selected candidate regions from a comparative genomics point of view, taking advantage of both finished and ongoing genome projects to study molecular divergence patterns in these regions between humans and other close primates.

c. Work plan

Timelines will proceed according to planned experiments. Exact timing will depend on results.

i) Phenotypic study.

i.1) Develop measures. **Within year 1.**

i.2) Perform laboratory experiments. **Within years 0.5 and 1.5**

ii) Genetic study.

ii.1) Genotyping. **Within years 1.5 and 2**

ii.2) Data analysis. **Within years 2 and 2.5**

ii.3) Functional analysis. **Within years 2 and 2.5**

ii.4) Comparative Study. **Within years 2 and 2.5**

Paper preparation and submission will be performed within the last year of the project.

d. Deliverables/milestones

The basic milestone of the project is to be able to uncover genetic differences between humans that may partly account for the diversity of human cooperative and trade-related behaviours. Another important result of our project will be the triggering of further research aiming to dissect the candidate regions or genes showing strongest association with the measured behaviours. Finally, results of the project will help to estimate the levels and structure of extant genetic variability in cooperative behaviour and, thus, would provide key information for the development of models as intended by other researchers within this project.

Specific deliverables will be:

- d.1) Development of consistent and stable behavioural measures that can be used in other studies.
- d.2) Performing behavioural studies on a group of >600 individuals.
- d.3) Detailed knowledge of the genomic distribution of associations with cooperative and trade-related behaviour.
- d.4) Knowledge of the genes or metabolic pathways involved in human behavioural variability.
- d.5) Detailed knowledge of the candidate genes or regions where causal variants might be detected.
- d.6) Potential to identify the recent action of natural selection upon human behaviours.
- d.7) Evaluation of the nature of human behavioural diversity in the context of evolution and genetics.
- d.8) Publication of results in high-impact, leading journals and presentation in international meetings.

e. Justification for budget items, especially staff costs and equipment

The presented budget contains two main items, staff costs and genotyping costs.

Staff costs. One postdoc and one student are requested. The postdoc will be an experienced and independent individual that will deal with both ends of the project: the preparation of proper behavioural measures and the design of experiments at one end and the final data analysis in the other end, specially as to copy number polymorphisms. The data analysis process, and particularly the functional and comparative analysis of the genes or genomic showing strong association with behavioural traits, is expected to be complex, so a postdoc with working knowledge in bioinformatics and statistical genomics will be needed.

The student will do his or her PhD within this project. He or she will be in charge of the sample gathering, preparation and management. The student will also contribute to the performance of behavioural experiments and to the analysis of the Whole Genome Scan, specially as to potential SNPs associated with behavioural traits. The student will be supervised by myself, with the assistance of the postdoc.

Genotyping costs. The cost for genotyping of 600 samples in Affymetrix's GeneChip Mapping 500k arrays, crucial to this study, is of €165,000.

e. Works cited

- Bouchard & Loehlin. 2001. Genes, evolution and personality. *Behavior Genetics* Vol 31, pp. 243-273.
- Brendgen et al. 2005. Examining Genetic and Environmental Effects on Social Aggression: A Study of 6-Year-Old Twins. *Child Development* vol 76, pp. 930-946.
- Cardon & Bell. 2001. Association study designs for complex diseases, *Nature Reviews* 2, pp. 91-99.
- Collins et al. 1999. Genetic epidemiology of single-nucleotide polymorphisms. *Proc. Natl Acad. Sci. USA* vol 96, pp. 15173-15177.
- Fisher. 1958. *The Genetical Theory of Natural Selection*, 2nd rev. ed., New York, Dover Publications.
- Hammock & Young 2005. Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science* vol 308, pp. 1630-1634.
- Harris. 1999. Review and methodological considerations in research on testosterone and aggression. *Aggression and Violent Behaviour* vol 4, pp. 273-291.
- Jiang & Huhtaniemi. 2004. Polymorphisms in androgen and estrogen receptor genes: effects on male aging. *Exp. Gerontology*. vol 39, pp. 1603-1611.
- Kosfeld et al. 2005. Oxytocin increases trust in humans. *Nature* vol 435, pp. 673-676.
- Lim et al. 2004. Enhanced partner preference in a promiscuous species by manipulating the expression of a single gene. *Nature* vol 429, pp. 754-757.
- Marchini et al. 2005. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature genet.* vol 37, pp. 413-417.
- Maynard Smith. 1982. *Evolution and the Theory of Games*. Cambridge University Press.
- Petronis. 2006. Epigenetics and twins: three variations on the theme. *Trends in Genetics*. vol 22 pp.347-350
- Risch. 2000. Searching for genetic determinants in the new millennium. *Nature* 405, pp. 847-856.
- Rot et al. 2006. Social behaviour and mood in everyday life: the effect of tryptophan in quarrelsome individuals. *J. Psychiatry Neurosci.* vol 31, pp. 253-262.
- Rushton. 2004. Genetic and environmental contributions to pro-social attitudes: a twin study of social responsibility. *Proc. Royal. Soc. Lond. Series B.* vol 271, pp. 2583-2585.
- Sharp et al. 2006a. Structural Variation of the Human Genome. *Annu. Rev. Genomics Hum. Genet.* vol 7, pp. 407-442.
- Sharp et al. 2006b. Segmental Duplications and Copy-Number Variation in the Human Genome. *Am. J. Hum. Genet.* vol 77 pp. 78-88.
- Smith et al. 2005. Factors affecting group spread within wild mixed-species troops of saddleback and mustached tamarins. *Int. J. Primatology* vol 26, pp. 337-355.
- Toga & Thompson 2005. Genetics of brain structure and intelligence. *Ann. Rev. Neuroscience.* vol. 28, pp. 1-23.

Tuvblad et al. 2006. The development of antisocial behaviour from childhood to adolescence. A longitudinal twin study. *Eur. Child Adolesc. Psychiatry* vol 24, pp. 216-225.

Van den Bogaert et al. 2006. Major Affective Disorders and Schizophrenia: A Common Molecular Signature? *Human Mutation* vol 27 pp. 833- 853.

Van Belzen & Heutink. 2006. Genetic analysis of psychiatric disorders in humans. *Genes, Brain & Behaviour* vol 5, pp. 25.

Zaboli et al. 2006. Tryptophan Hydroxylase-1 Gene Variants Associate with a Group of Suicidal Borderline Women. *Neuropsychopharmacology* vol 31, pp. 1982-1990.

Section C

API: The Behavioural Genetics of Risk, Time, and Social Preferences

Ernst Fehr

University of Zurich

C.1. AP Contribution to the CRP

a. Associated Project aims and objectives

The contribution of the group working around Professor Ernst Fehr of the University of Zurich to the TECT project covers the following areas:

Genetic factors appear to play a decisive role in behaviour. We have evidence from a Swedish team around Magnus Johannesson, who conducted experiments with heterozygous and homozygous twins, that the correlation of behaviours among homozygous twins is much larger, suggesting a significant role for genetic factors in risk, time and social preferences.

Our research group will collaborate with Arcadi Navarro of the University of Pompeu Fabra in Barcelona, Spain, whose interests also cover the genetic and evolutionary backgrounds of basic human behaviours.

b. Methodologies/experiments

Much work in behavioural genetics is not really based on observed behaviour but on reported behaviour, i.e. on behaviour reported in questionnaires. In contrast to this approach, we intend to measure risk, time and social preferences directly in laboratory experiments. Experimental economics and experimental psychology provides a rich and fertile menu of behavioural tasks, enabling us to measure this important dimensions of behaviour.

We plan to run experiments with about 600 people. We will take saliva samples from them to determine associations between genes and behaviours. We will also include a subsample of heterozygous and homozygous twins in our experiments.

We will design and conduct the experiments at the University of Zurich where we also have an expert in behavioural genetics in our team. The whole project will be executed in close collaboration with Arcadi Navarro of the University of Pompeu Fabra in Barcelona, Spain

c. Work plan

Our first task is to define stable behavioural phenotypes in terms of risk, time and social preferences. This task will be solved using behavioural experiments that measure a variety of behaviours of the same subjects across time and across different experiments. Then, after we have assessed the stable phenotypes, we run experiments on risk, time, and social preferences with 600 subjects. We will analyze the DNA of 300 subjects on a large scale. Based on the results from this sample – in terms of correlations between specific genes or gene clusters with the measured behaviours – we predict that similar correlations will hold in the whole sample and in the second part of the sample. If this turns out to be true, we will have strong evidence that the observed correlations are not spurious. Finally, based on the specific genetic findings – e.g., an involvement of the dopaminergic and the serotonergic system in risk, time, or social preferences – we plan to conduct pharmacological experiments with the relevant agonists and antagonists. These experiments constitute the last step in this research program.

d. Deliverables/milestones

Depending on how fast we can establish the stable phenotypes, we expect to accumulate data and have presentable results within three years. We expect several top-notch publications in leading journals to result from this project.

AP2: Metacognition and decision-making under risk: individuals and groups***Alex Kacelnik******Oxford University*****B.1. Associated Project contribution to the CRP****a. AP aims and objectives**

The Behavioural Ecology Research Group at Oxford University led by Alex Kacelnik will focus on theoretical and empirical analysis of individual and group decision-making under risk, with an emphasis on the role of subject-generated uncertainty.

This project relates to the proposal by Gintis' framework for the unification of behavioural sciences, and it includes theory from economics, evolutionary biology, behavioural ecology and psychology and because it deals with empirical research in human and animal models. It also relates to the project proposed by Fehr because in the latter one of the central components is to establish reliable behavioural phenotypes in terms of risk, and this can be aided by considerations of the many dimensions of risky behaviour as discussed above. Another close link is with the work proposed by Mace, because the putative differences in prosocial behaviour between cultural units or societies may be located at the point where knowledge is shared, and our software is ideally designed for measuring these variables. Indeed, Mace's Ph D student Shakti Lamba was, as an M Sc student in Oxford, responsible for generating the pilot work on the behaviour of dyads that we quoted above.

Dealing with metacognition in the context of risk has the attraction of being strongly multidisciplinary: ideas from economics, evolutionary theory, behavioural ecology, and psychology contribute to define the problem and to generate the experimental approach. Adding the supra-individual dimension associated the problem with the functional analysis of cooperative behaviour and group functioning.

b. Methodologies/experiments

Risk-related research varies according to the source and nature of unpredictability. The most widespread approach deals with "*Risk*" per se, and concerns strategies to make choices in the presence of stochasticity in the world (an example is gambling, and a biological counterpart is the tuning of bird migration to the conditions at a far off destination target). Orthodox economics, Prospect Theory and Risk Sensitivity Theory address these issues from the viewpoint of economics, cognitive psychology and evolutionary biology respectively (Kahneman and Tversky 1979; Mas-Colell, Whinston et al. 1995; Kacelnik and Bateson 1997). A second issue is "*Uncertainty*" where unpredictability depends on

incomplete knowledge, as in medical decisions involving new drugs or animal foraging where feeding rate can be improved by better sampling the local habitat (Krebs, Kacelnik et al. 1978). Risk and uncertainty are formally different (as only the latter deals with investment in knowledge acquisition) but merge into each other at some time scales (viz. with extremely short time horizons as when facing a one-shot choice).

There is a third sort of problem, tackled by Scalar Expectancy Theory, (Gibbon 1977; Reborada and Kacelnik 1991; Kacelnik and Brito e Abreu 1998) in which the unpredictability is irreducible as in risk but depends on the subject's own information-processing system. Here there is some level of error even when the world is fully known and deterministic, because variance is generated by perception and/or handling of information. For instance, when optimal behaviour requires comparing time intervals, the limited accuracy of the internal clock transforms the problem into one of risk (Kacelnik and Brunner 2002). This problem belongs in the risk category because it assumes full knowledge and irreducible unpredictability, but it also relates to uncertainty because the source of unpredictability is internal. Similar issues are treated within Bounded Rationality (Simon 1955; Gigerenzer 2004) because there the subject includes in its decision information about its own limitations and in Prospect theory because subjects may actually have distorted perceptions of objective probabilities but somehow neutralise these distortions by biasing their decision thresholds (Kahneman and Lovallo 1993).

c. Work plan

This project will target an issue that straddles across all of the categories described above. As an illustration, imagine that a student chooses between taking a voluntary oral examination or accepting the grading achieved so far by coursework. Taking the exam may lead to marks that are higher or lower than those secured at present. An optimal decision-maker would be sensitive to the usual variables of risk analysis (namely, the shape of the utility function and the present state or point of reference—viz. whether the present mark is sufficient to qualify for PhD acceptance). In addition, however, optimal subjects would judge their confidence in their own self-assessment, as this influences the mean and variance of the expected reward.

Problems of this kind, where the subject needs to assess her own knowledge having dealt with under the label of Metacognition (Flavell 1979). Although initially the ability of thinking about cognitive dimensions was assumed to be an exclusively human feature, decision-making research involving metacognition acquired new momentum when researchers brought the problem to non-human subjects (Inman and Shettleworth 1999; Hampton 2001; Smith, Shields et al. 2003). Dealing with non-human subjects makes a big difference because it calls for methods of assessing metacognitive information

without self-reporting, and without extrapolating from observed choices (as the goal is to establish the relation between both). Inspired by experimental protocols developed in animal research on metacognition, here we return to humans, integrating metacognition into issues related to risky decision-making.

d. Deliverables/milestones

To perform the above tasks, we have developed a software instrument (APES) that objectively dissociates subjects' performance in a psychophysical discrimination from their judgements about this performance. Experiments designed around this package allow for testing theoretical models that incorporate the degree of confidence in metacognitive assessments. We will examine how extant theories of risk sensitivity cope with stochasticity generated by imprecise self-knowledge and propose new theoretical modifications to account for the obtained results.

A further dimension of the problem refers to supra-individual decision-making. In many situations (for instance grant allocation by committees) groups combine knowledge and utility functions to take joint decisions. In doing this, there is a collective cooperative evaluation and weighting of group members' expertise, and individual members (ideally) regulate the forcefulness of their arguments according to metacognitive judgements. Our intention is to quantify this process by employing our software in experiments involving groups and in situations where several subjects contribute knowledge and one decides (many-to-one), as well as situations where one subject produces information and several reach a decision (one-to-many). We have run pilot experiments using psychophysical tasks implemented in APES with individuals and dyads. In addition to finding age and gender effects, we found that dyads had lower error in their self-assessment, and achieved higher payoffs than individuals. Sharing information and decision power is likely to have been a major issue in the development of cooperation in animal and human groups. We propose to examine rather precisely how this sharing operates.

e. Works cited

- Flavell, J. H. (1979). "Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry." *American Psychologist* **34**: 906-911.
- Gibbon, J. (1977). "Scalar expectancy theory and Weber's law in animal timing." *Psychological Review* **84**: 279-325.
- Gigerenzer, G. (2004). Striking a blow for sanity in theories of rationality. *Models of man: Essays in honor of Herbert Simon*. B. Augier and J. G. March. Cambridge, MA., MIT Press: 389-409.
- Hampton, R. R. (2001). "Rhesus monkeys know when they remember." *Proceedings of the National Academy of Sciences, USA* **98**(9): 5359-5362.

- Inman, A. and S. J. Shettleworth (1999). "Detecting metamemory in nonverbal subjects: A test with pigeons." Journal of Experimental Psychology: Animal Behaviour Processes **25**(3): 389-395.
- Kacelnik, A. and M. Bateson (1997). "Risk-sensitivity: cross-roads for theories of decision-making." Trends in Cognitive Sciences **1**(8): 304-309.
- Kacelnik, A. and F. Brito e Abreu (1998). "Risky choice and Weber's Law." Journal of Theoretical Biology **194**(2): 289-298.
- Kacelnik, A. and D. Brunner (2002). "Timing and foraging: Gibbon's scalar expectancy theory and optimal patch exploitation." Learning and Motivation **33**: 175-195.
- Kahneman, D. and d. Lovallo (1993). "Timid choices and bold forecasts: a cognitive perspective on risk taking." Management Science **39**: 17--31.
- Kahneman, D. and A. Tversky (1979). "Prospect Theory: An analysis of decision under risk." Econometrica **47**: 263-291.
- Krebs, J. R., A. Kacelnik, et al. (1978). "Tests of optimal sampling by foraging great tits." Nature **275**: 27-31.
- Mas-Colell, A., M. D. Whinston, et al. (1995). Microeconomic Theory. Oxford, OUP.
- Reboreda, J. C. and A. Kacelnik (1991). "Risk sensitivity in starlings: variability in food amount and food delay." Behavioural Ecology **2**: 301-308.
- Simon, H. A. (1955). "A behavioural model of rational choice." Quarterly Journal of Economics **69**: 174-183.
- Smith, J. D., W. E. Shields, et al. (2003). "The comparative psychology of uncertainty monitoring and metacognition." Behavioural and Brain Sciences **26**: 317-373.

AP3: The development of co-operative norms: an empirical investigation

Ruth Mace

University College, London

B.1. Associated Project contribution to the CRP

a. AP aims and objectives

Frequent and large-scale cooperation between genetically unrelated individuals makes human societies a great anomaly in nature (Fehr & Fischbacher, 2003). Numerous studies in many parts of the world have demonstrated that individuals do not always seek to maximise material gain (Roth *et al* 1991, Charness & Rabin 2002, Henrich *et al*, 2001). Human beings care about fairness norms and even enforce them at a cost to themselves (Fehr & Fischbacher, 2004). Cooperation in human societies is largely based on social norms. The existence of such norms is one of the biggest unsolved evolutionary puzzles. How does unselfish behaviour evolve and persist in a population where selfish defectors would be more successful? The current challenge is to explain the initiation, spread and maintenance of unselfish, prosocial norms (Ehrlich & Levin, 2005).

While there is no dearth of theoretical models addressing the evolution of prosocial behaviour in humans, the behaviour of real populations is still poorly understood. A cross-cultural games project in 15 small-scale societies (Henrich *et al*, 2004) has demonstrated that there are significant differences in normative behaviour between cultural groups. The payoffs to cooperation, the level of market integration and other institutional features have high predictive power in explaining this variability between ethnic groups (Henrich *et al*, 2004). However, variation within cultural groups has not been explained. This variation within cultural groups provides a unique opportunity to explore the very origins of prosocial norms, through an examination of both group characteristics and environmental variables, while controlling for larger cultural differences.

We wish to attempt an analysis of the ‘ecological’ variation between communities belonging to the same cultural group, in order to tease apart the extent to which individual or population level differences influence the development of prosocial norms. Through these analyses we will also attempt to throw light on the mechanisms of cultural evolution – which help assess the ecological plausibility of models developed in this project especially by Bowles and Rowthorn on niche construction, Castelfranchi on norm violation and Kerszberg on the development of social norms.

Key questions are:

Do properties associated with social organization (e.g. distribution and accessibility of resources, degree of immigration/emigration, population size) affect the norms adopted by a group? The effect of group size is of key interest as many evolutionary models assume that large group size inevitably weakens co-operative behaviour whereas many economists do not make that prediction and one model (Guzman et al in press) predicts co-operation can emerge more easily in large groups. Further, even moderate rates of migration between groups are usually predicted to disrupt the evolution of co-operative norms according to most evolutionary models. How prosocial behaviour is actually influenced by these variables in real human populations has barely begun to be investigated. What little evidence there is suggests large groupings may show more prosocial behaviour than small groups (eg Marlowe 2004), but that is a matter for further investigation. We aim to identify how group size and migration patterns influences co-operative norms in an Indian population comprising many sub-communities of varying size.

Does the behaviour of individuals in economic games reflect a dispositional or context specific strategy? Individuals have different costs and benefits associated with co-operation that are likely to be influenced by such variables as their relatedness to the rest of the group. We will examine determinants of individual variation responses to ultimatum and other games to determine the relative importance of group properties discussed above versus individual costs and benefits.

What are the processes that drive ‘cultural evolution’ of prosocial behaviour? What are the time scales over which they occur? New immigrants into a group, such as women moving into the community to marry, may behave according to norms of their natal village or their new village, or this may be a function of the length of time in which they have inhabited the new village. First and second generation immigrants into a community will also be compared for temporal differences in prosocial norms.

Does the degree of cooperative clustering in a social group affect networking and information flow, which may in turn feedback into the nature and degree of cooperation? Network analyses will be used to identify important attributes of the social structure of communities and whether these account for any of the observed variation between communities.

b. Methodology/experiments

We will use economic games, primarily the ultimatum game and public goods games, to measure prosocial behaviour in different communities belonging to the same ethnic group. The study site has been chosen to contain communities varying in population size, migration rates, distance from a major

town or water source and other ecological parameters. It contains women who marry within their natal village as well as women who moved to their current village to marry. Thus the variation in responses in ultimatum games and public goods games can be analysed in a way that will enable both individual level and group level differences to be identified and their relative importance determined.

The research will be conducted amongst the Gond of central India, found in the states of Maharashtra, Madhya Pradesh and Chhattisgarh. The Gonds are one of the largest tribal groups of India and much is already known about their cultural history (Das 1979, Fürer-Haimendorf 1981). Now primarily agriculturists, the geographical variation of this endogamous group makes it ideal for the kind of ecological analysis of prosocial behaviour that we wish to attempt. Women marry both within and between villages. While the tribe has an indigenous language ('Gondi') most villages are also Hindi speaking. The researcher's knowledge of Hindi will provide easier access in conducting economic games.

In line with established practices, games will be paid for real money for non-trivial sums, usually equivalent to around one or two days pay for the individuals concerned.

c. Work plan

The research will be conducted by UCL PhD student Shakti Lamba, in collaboration with her supervisor Ruth Mace. SL will spend the majority of year 1 in the UK, but making contact with other project partners including Sam Bowles (SFI) and Rob Rowthorn (Cambridge) and Herb Gintis (Hungary), whose various models make direct predictions about how group dynamics influence co-operative norms. A pilot study will be undertaken in India to establish effective experimental protocols.

In year 2 SL will spend the majority of her time in India conducting economic games throughout the diverse range of communities in the study site.

In year 3 data will be analysed primarily at UCL, but again with visit to other partners listed above.

d. Deliverables/milestones

Year 1: Permissions obtained, precise fieldwork methods piloted and full experimental protocol devised, in discussion with other consortium members

Year 2: Data collected on ecological and individual variation in behaviour in economic games completed.

Year 3: Analysis completed on individual and group level determinants of prosocial norms, and how these results support or invalidate assumptions made in the various models of the evolution and development of co-operative norms.

e. Works cited

Charness, G. & Rabin, M. (2002) [Understanding Social Preferences With Simple Tests](#). *The Quarterly Journal of Economics* 117(3) 817-869

Das, N. C. (1979) Demographic aspects of the Gonds, 1961-71. *Man in India* 59(3) 218-34

Ehrlich, P.R. & Levin, S. A. (2005) The evolution of norms. *PLoS Biology* 3(6) 943- 948

Fehr, E. & Fischbacher, U. (2003) The nature of human altruism. *Nature* 425, 785-791

Fehr, E. & Fischbacher, U. (2004) Social norms and human cooperation. *Trends in Cognitive Sciences* 8(4) 185-190

Fürer-Haimendorf, C. (1981) *The Gonds of Andhra Pradesh*. London: George Allen & Unwin

Guzman R, Rodriguez-Sickert C & R Rowthorn (in press) When in Rome, do as the Romans do: The co-evolution of altruistic punishment, conformist learning, and cooperation. *Evolution and Human Behaviour* (in press).

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R. (2001) In search of homo economicus: experiments in 15 small scale societies. *American Economic Review*, 91(2) 73-78

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. (2004) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press

Roth, A. E. *et al* (1991) Bargaining and market behaviour in Jerusalem, Ljubljana, Pittsburgh and Tokyo: an experimental study. *The American Economic Review* 81 (5) 1068-1095

B.3 Related Projects (information on funding)

The major part of the funding for this project has been confirmed from the Cogito Foundation and more will be sought in year 1 from the WennerGren Foundation and other sources yet to be determined, to supplement the costs of fieldwork including economic games in year 2, and for visits to consortium partners.

AP4: Evolution of moral behaviour

Robert Rowthorn

Faculty of Economics, Cambridge UK

B.1. Associated Project contribution to the CRP

a. AP aims and objectives

I first became interested in the evolution of moral behaviour when I participated in the Biology & Economics group at King's College, Cambridge in the mid-1990s. During that period, I wrote a paper that was eventually published some years later after being revised following a visit to the Santa Fe Institute in 2005. I also began a long-term collaboration with Carlos Rodríguez-Sickert and his brilliant PhD student, Ricardo Guzmán. The work that we are doing is directly related to work which has been recently published by other members of TECT group (Bowles, Boyd, Gintis, Richerson). Moreover, both Rodríguez-Sickert and I are linked to the Santa Fe Institute under the aegis of Sam Bowles.

Related papers that the three of us have written in the near past: (1) Rowthorn, R.(2006), "The evolution of altruism between siblings: Hamilton's rule revisited", *Journal of Theoretical Biology* 241(4): 774-90. (2) Guzmán, R., C. Rodríguez-Sickert and R. Rowthorn (2006), "When in Rome, do as the Romans: The coevolution of punishment, conformism and cooperation", *Evolution and Human Behaviour*, forthcoming. (3) Rodríguez-Sickert, C., R. Guzmán & J.C. Cárdenas (2006). "Institutions Influence Preferences: Evidence from a Common Pool Resource Experiment", revised and resubmitted to *Journal of Economic Behaviour and Organization*. (4) Kuzmics, C. A. and C. Rodríguez-Sickert "The Evolution of Moral Codes of Behaviour", submitted to *American Economic Review*.

b. Methodologies/experiments

Our research agenda has two components: one theoretical and the other experimental.

I. Theoretical Component.

Alternative Structures of Social Interaction

Our intention is to investigate the interaction between learning strategies and social behaviour within the context of different forms of social interaction for those considered in the EH&B article, for example, coordination or distribution problems.

Group Size & Hierarchies

In current models, group size is exogenous. We do not believe that the puzzle of cooperation in large groups will be solved until group size itself is allowed to co-evolve along with cooperation. People get together for a reason: to seize the gains of cooperation and specialization. Large groups benefit from scale economies. They command larger armies than small groups and allow greater degrees of specialization. However, there are forces operating against larger groups. The increasing advantage of free riding (due to monitoring problems) is one such force. Another is increasing difficulty of coordination in a large society. Is the emergence of hierarchies a solution that human societies have found to cope with these problems?

Migration and assimilation.

Migration gives rise to problems which are often similar to those large groups size. Benefits deriving from cultural diversity, such as innovation, may be outweighed by communication difficulties and parochial loyalties that obstruct group level collective action. For this reason, inter-group competition, may confer an advantage upon groups that limit diversity by severely restricting immigration from other groups or else by promoting the rapid assimilation of immigrants. The evolution of such social mechanisms may have genetic counterparts in the form of a high propensity to conform or a high propensity towards xenophobia.

II. Experimental Component.

There is a growing experimental body of evidence (e.g., Efferson et al. 2006) in which it is explored how alternative forms of social learning are used to obtain information from an environment detached from social interaction. However, how these mechanisms co-evolve in an environment in which individual decisions involve strategic interaction with others has not been explored. We intend to set up experimental micro-societies that play the societal game as in our E&HB article under different information conditions. The information conditions we could test are:

- Information about role model payoff: At the end of every round each player is informed about the payoff of a randomly selected role model.
- Information about behavioural frequencies: At the end of every round all players are informed of the frequencies of each kind of behaviour.

In both types of condition, players will always know their own payoffs.

The two information conditions give rise to four experimental treatments:

Case #1: Information about role model payoff + information about behavioural frequencies;

Case #2: Information about role model payoff + NO information about behavioural frequencies;

Case #3: NO information about role model payoff + information about behavioural frequencies;

Case #4: NO information about role model payoff + NO information about behavioural frequencies;

In case 4, only individual learning is possible. In all other cases, social learning may take place.

We fit the experimental results to a learning model that incorporates the possibility of both social and individual learning. From the fitted model we should be able to infer which kind of learning rules (or combination of learning rules) is used by modern humans in each setting.

c. Work plan

Our joint research is concerned with the relationship between behavioural transmission mechanisms and the emergence pro-social behaviour. This is the topic of our forthcoming E&HB article, and our research agenda is to extend the analysis of that paper in various directions. In our E&HB paper, we model the co-evolution of behavioural strategies and social learning rules in the context of a cooperative dilemma, a situation in which individuals must decide whether or not to subordinate their own interests to those of the group. There are two learning rules in our model, conformism and payoff-dependent imitation, which evolve by natural selection; and three behavioural strategies, cooperate, defect, and cooperate and punish defectors, which evolve under the influence of the prevailing learning rules. Group and individual level selective pressures drive evolution. We also simulate our model for conditions that approximate those in which early hominids lived. We find that conformism can evolve when the only problem that individuals face is a cooperative dilemma, in which pro-social behaviour is always costly to the individual. Furthermore, the presence of conformists dramatically increases the group size for which cooperation can be sustained. The results of our model are robust: they hold even when migration rates are high, and when conflict among groups is infrequent.

Annex 1

Herbert Gintis

Degrees

Ph.D. in Economics, Harvard University, 1969

Position

Professor, 2005—present

Organisation: Central European University
Department of Economics
1051 Budapest, Nador u. 9.
Email: gintish@ceu.hu

Website: www.ceu.hu

Grants

Co-director, with Professor Robert Boyd (Anthropology, UCLA), of "Economic Environments and the Evolution of Norms and Preferences," a multi-year, interdisciplinary research project funded by the John D. and Catherine T. MacArthur Foundation, 1997--2006.

Publications

Game Theory Evolving (Princeton: Princeton University Press, June 2000)

with Samuel Bowles and Melissa Osborne, *Unequal Chances: Family Background and Economic Success* (Princeton University Press, 2005)

Herbert Gintis, "Towards the Unification of the Behavioral Sciences", Behavioral and Brain Sciences (2006).

with Samuel Bowles, Robert Boyd, and Ernst Fehr, **Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life** (Cambridge, MIT Press, 2005).

with Joe Henrich, Robert Boyd, Samuel Bowles, Colin Camerer and Ernst Fehr, **Foundations of Human Sociality: Ethnography and Experiments in Fifteen Small-scale Societies** (Oxford: Oxford University Press, 2004).

SAMUEL BOWLES

Education:

B.A., Yale University, 1960.

Ph.D. (Economics), Harvard University, 1965.

Principal Positions Held:

Harvard University, Associate Professor of Economics (1971-74) and Assistant Professor of Economics (1965-71).

University of Massachusetts at Amherst, Professor of Economics, 1974 to present (now Emeritus).

Santa Fe Institute, Arthur Spiegel Research Professor and Director of the Behavioral Sciences Program (2003 B) Previously, Director of the Economics Program 2000-2003.

Research Network on Inequality and Economic Performance (MacArthur Foundation), Founder and Co-director, 1993-present.

University of Siena, Professor, Faculty of Economics, 2003-present

Organisation:

Santa Fe Institute and University of Sienna,

1399 Hyde Park Road

Santa Fe, NM 87501, USA

email: bowles@santafe.edu

<http://www.santafe.edu/~bowles>

Prof. Eörs Szathmáry**Date of Birth:** 18-Dec.-1959**Place of Birth:** Budapest, HUNGARY**Nationality:** Hungarian**Postal address:** H-1014 Budapest, Szentháromság u. 2**Phone:** +36-1-2248300**Fax:** +36-1-2248310**E-mail:** szathmary@colbud.hu**Positions**

- 1995 – present Permanent fellow, Collegium Budapest (Institute for Advanced Study)
- 1997 - present Professor of biology, Dept of Plant taxonomy and Ecology, Eötvös University, Budapest
- 1995-2002 Head of department and Ph. D. programme in theoretical biology and ecology, Dept of Plant taxonomy and Ecology, Eötvös University, Budapest
- 1995 D. Sc. (Doctor of the Hungarian Academy)
- 1995 Associate professor, Dept of Plant taxonomy and Ecology, Eötvös University, Budapest
- 1994-1995 Fellow, Collegium Budapest (Institute for Advanced Study)
- 1994 Guest professor, Institute of Zoology, University of Zürich, Switzerland
- 1992-1993 Fellow, Institute for Advanced Study Berlin
- 1991-1992 Research fellow, Laboratory of Mathematical Biology, National Institute of Medical Research, London
- 1987-1995 Research fellow, Dept of Plant taxonomy and Ecology, Eötvös University, Budapest
- 1987-1988 Postdoctoral scholarship with John Maynard Smith, The University of Sussex, Brighton

Education

- 1984 Ms.S., Biology, Eötvös University, Budapest.
Thesis: *Dynamical coexistence in prebiotic systems*
- 1987 Ph.D, Ecology, Eötvös University, Budapest.
Thesis: *The roots of individual organization*

Honours

- 1995 Outstanding teacher of the Faculty of Natural Sciences, Eötvös University, Budapest
- 1996 New Europe Prize for Higher Education and Research, Stanford
- 1999 Prize of the Academy, Hungarian Academy of Science, Budapest
- 2001 Member of Academia Europaea

Cristiano Castelfranchi

Date of Birth: 08 June 1944

Place of Birth: Rome, Italy

Nationality: Italian

Positions

Full Professor of Cognitive Sciences at the University of Siena and Director of the interdisciplinary Institute of Cognitive Sciences and Technologies of the National Research Council (ISTC), in Rome. He was formerly Head of the Division of Artificial Intelligence, Cognitive and Interaction Modelling, and of the Division of Social Psychology at the same institute. He has been Principal Investigator in the CRP “The cultural self-organization of cognitive grammar” within the Eurocores Programme OMLL (ended in 2004) and of the CRP “Consciousness in Interaction” within the EUROCORES Programme CCNC. He has also been involved in many European Projects and Networks under the IV, V and VI Frameworks (e.g. ALFEBIITE, HUMAINE and MINDRACES).

Organisation: Institute of Cognitive Sciences and Technologies CNR

viale Marx 15 - 00137 Roma, Italy , Website: <http://www.istc.cnr.it>

Telephone: +39-06-86090235

Email: cristiano.castelfranchi@istc.cnr.it

Research

His research activity is within the field of Cognitive Science and it has been mainly focused on the development of models of cognitive architectures and mental representations both for human and artificial agents. A constant focus of his research has been the representations of beliefs and goals, their interrelations and dynamics, their role in social behavior (from language to cooperation and social exchange) and how beliefs and goals mediate the emergence of macro-social phenomena and the control of society over the individuals. Of particular relevance for this project is the application of the goals theory to social interaction with a special emphasis on the role of evaluations and values, of social emotions (guilt, shame, pity, trust etc.), of power and dependence as objective structures and on the cognitive representation of norms. In his work, Castelfranchi has applied the conceptual and methodological tools of various disciplines (social psychology, clinical psychology, sociology and distributed artificial intelligence) within the cognitive science framework beyond current boundaries. Cristiano Castelfranchi is author of 16 books (5 in English and 2 in preparation for MIT and Wiley & Sons, 11 in Italian) and more than 300 articles in English and Italian, published on international journals, edited volumes and peer-reviewed proceedings. He is also a member of the editorial board of the international journals Autonomous Agents and Multi-Agent Systems and Cognitive Science Quarterly, and of CogNet at the MIT, Boston. Cristiano Castelfranchi has been either scientific chair, local organizer or invited speakers in several world-leading scientific meetings and conference in the field of knowledge dynamics, artificial intelligence, social simulation, and psychological dynamics.

Michel Kerszberg

Born 1951

Engineer (Brussels Free University) 1975

Ph. D. in **physics** (Weizmann Institute, Israel) 1980

Postdocs at Harvard University Physics dept (1980-1982) and at the **Xerox Palo Alto Research Center** (PARC, 1982-1984)

Worked for 10 years in the Molecular Neurobiology Unit at the **Pasteur Institute**. Currently head of the **Biological Modeling Unit** of UMR 7138, “Systématique, Adaptation, Evolution” at Université Pierre et Marie Curie (Paris 6).

Most relevant to the project, Michel Kerszberg has worked extensively with neural network models, both gene an learning-based; on population dynamics; and he has developed the *ctrl-Dev* application, which embodies an agent-based approach to molecular and cellular biological systems, including possible applications in neural development. MK has extensive experience with Unix/X-Windows GUI programming and C/C++ scientific programming. He collaborates with several groups in neurobiology and cellular and developmental biology, whose work is also of great relevance to the project

Arcadi Navarro

Title: Dr.

Name: Arcadio Navarro Cuartiellas

Passport number: 34734224.

Birth date: August the 13th 1969.

Home Address: C/ Sant Quintí 70. Entl 1a. 08041 Barcelona.

Professional Address: Departament de Ciències Experimentals i de la Salut.

Universitat Pompeu Fabra.

C/ Dr. Aiguader, 80.

08003, Barcelona.

Spain

Phone: +34 93 5322116 (Home) / +34 93 5422841 (University)

FAX: +34 93 5422802

E-MAIL: arcadi.navarro@upf.edu

WEB: <http://www.upf.edu/cexs/recerca/bioevo/CV-ArcadiNavarro.htm>

FUNDING AGENCY: MEC

Dr. Arcadi Navarro is professor of evolution and researcher at the Faculty of Life Sciences of the Universitat Pompeu Fabra in Barcelona. He is an evolutionary human geneticist with skills in theoretical population genetics, computational genomics and molecular evolution. He is also the Bioinformatics Coordination in Spain's National Genotyping Center (<http://www.cegen.org>). He has worked in human evolution from the genomic and the theoretical perspectives.

EDUCATION

PhD in Genetics. Grade: *Cum laude*. Universitat Autònoma de Barcelona (U.A.B), Spain,

MSc in Genetics. Grade: Excellent. Universitat Autònoma de Barcelona (U.A.B). Spain.

BSc in Biology. Universitat Autònoma de Barcelona (U.A.B). Spain.

SELECTED AWARDS AND HONORS

2005. Eleventh European Scientific Divulagation Award (2nd prize). Awarded by the Universitat de València.

2004. TOYP Prize. (The Outstanding Young Person) in the category of Biomedical Research. Awarded by the Junior Chambers of Catalonia.

1992. Best Student Price. Awarded by the Universitat Autònoma de Barcelona.

Ernst Fehr

Position:

Director, Institute for Empirical Research in Economics, University of Zürich, Switzerland

Organisation:

Institut für Empirische Wirtschaftsforschung
Universität Zürich, Blümlisalpstrasse 10
CH - 8044 Zürich

Date of Birth: 21. June 1956

Homepage: <http://www.iew.unizh.ch/home/fehr/>

Awards and Honors:

President, Economic Science Association, 2003-2005.
Honorary Doctorate, University of St. Gallen

ALEJANDRO KACELNIK

Department of Zoology, Oxford University, Oxford OX1 3PS UK

Tel: +44 1865 271164, FAX: +44 1865 271120

Web site: <http://users.ox.ac.uk/~kgroup>

Biographical details Born 14 December 1946, Buenos Aires, Argentina. Dual British-Argentinian nationality. Married, one child.

Current position

Professor of Behavioural Ecology, Department of Zoology, Oxford University. And Vice-Gerent, Pembroke College, University of Oxford

E P Abraham Research Fellow, Pembroke College, Oxford.

Education

DPhil (Zoology, Animal Behaviour), Oxford. (Supervisor: J R Krebs)

Licenciado en Ciencias Biologicas (Major: Ecophysiology), University of Buenos Aires.

Visiting Professorships/Distinctions

Graduate Award, Wolfson College.; Van der Klaauw Professor, Leiden; .Visiting Professor at Indiana University (Bloomington);, Université Claude Bernard, Lyon, France., University of Buenos Aires, Princeton University. Member of the Academia Europea; Co-winner (with E. Fehr) of the Cogito Prize for interdisciplinary research. Fellow, Wissenschaftskolleg zu Berlin. Moderator of Focus Group: The Sciences of Risk.

Editorial Boards and Memberships

BMC- Ecology; Learning & Behavior [ex-Animal Learning and Behaviour] (Consulting editor 2003-onwards); Ethology (Associate editor 1999-2001); Behavioral Ecology and Sociobiology (Associate Editor 1997-2001).; The American Naturalist (Associate Editor 1991-1995).; Quarterly Journal of Experimental Psychology (Associate Editor 1990-1993).; Behavioural Processes (1990-); Behavioural Ecology (1990-96).; Symposia on Quantitative Analyses of Behavior (1986- 1990). International Society for Behavioral Ecology; British Ecological Society; Experimental Psychology Society Society for Quantitative Analyses of Behavior; Psychonomics Society; Association for the Study of Animal Behaviour. Ramon Y Cajal Fellowship Ctee (Spain); BBSRC Animal Sciences Panel; Royal Society URF Committee.

Invited lectures and departmental seminars

Approximately 260 invited lectures and conference presentations, including the following:

RUTH MACE

Work address: Department of Anthropology, University College London,
14 Taviton St, London WC1H 0BW, UK

tel. +20 7679 8845

email: r.mace@ucl.ac.uk

Current position: Professor of Evolutionary Anthropology

EDUCATION AND EMPLOYMENT HISTORY

Wadham College, Oxford (undergrad and D.Phil)	1980 -1987
Renewable Resources Assessment , Imperial College, London	1987 - 1989
School of Development Studies, University of East Anglia	1989 - 1991
Department of Anthropology, University College London	1991 - present

EDITORSHIPS

Editor: Evolution and Human Behaviour Associate editor: Proceedings of the Royal Society of London:
B, & Human Nature

PRIZES

1987 The Thomas Henry Huxley Medal awarded by the Zoological Society
of London for the best doctoral thesis submitted in 1987

The Gabriel Lasker Prize awarded by the journal Human Biology for the most significant
paper published in Human Biology in 1997

2003 The Curl Lecture 2003. Royal Anthropological Institute.

ROBERT ROWTHORN:

Work

Address: Economics Faculty, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom

Telephone: (44) 1223-335230

Fax: (44) 1223-

E-mail: rer3@econ.cam.ac.uk

Home

Address: 9 Willis Road, Cambridge CB1 2AQ, United Kingdom

Telephone: (44) 1223-368235

Academic Posts

Professor of Economics, University of Cambridge, UK

Fellow of Kings College, Cambridge, UK

Relevant Teaching

MPhil course on the economics of institutions (covers the evolution of institutions and human behaviour); the economics of natural resources.

Relevant Research

The evolution of institutions and human behaviour, including altruism and cooperation; economic aspects of species extinction; economics aspects of epidemics.

Annex 2

PI1: Herbert Gintis

Game Theory Evolving (Princeton: Princeton University Press, June 2000)

with Samuel Bowles and Melissa Osborne, *Unequal Chances: Family Background and Economic Success* (Princeton University Press, 2005)

Herbert Gintis, "Towards the Unification of the Behavioral Sciences", *Behavioral and Brain Sciences* (2006).

with Samuel Bowles, Robert Boyd, and Ernst Fehr, **Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life** (Cambridge, MIT Press, 2005).

with Joe Henrich, Robert Boyd, Samuel Bowles, Colin Camerer and Ernst Fehr, **Foundations of Human Sociality: Ethnography and Experiments in Fifteen Small-scale Societies** (Oxford: Oxford University Press, 2004).

PI2: Samuel Bowles

Bowles, S. & Posel, D. Genetic relatedness predicts South African migrant workers' remittances to their families. *Nature* **434**, 380-383 (2005).

Henrich, Joe, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, (et al) 2005. "'Economic Man' in Cross-Cultural Perspective: Behavioral experiments in 15 small-scale societies." *Behavioral and Brain Sciences*, 28.

Bowles, S. & Gintis, H. Persistent parochialism: trust and exclusion in ethnic networks. *Journal of Economic Behavior & Organization* **55**, 1-23 (2004).

Bowles, S. & Gintis, H. The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical Population Biology* **65**, 17-28 (2004).

Bowles, S., Choi, J. K. & Hopfensitz, A. The co-evolution of individual behaviors and social institutions. *Journal of Theoretical Biology* **223**, 135-147 (2003).

PI3: Eörs Szathmáry

(full list is available at http://www.colbud.hu/fellows/szathmary_pub.html)

1. Gulyás, B. & Szathmáry, E. (2002) Monkeys – a great asset to reveal human cognitive functions. *Neuroreport* **13**, 2167-2168.

2. Szathmáry, E. (2002) Cultural processes: the latest major transition in evolution. In: L. Nadel (ed.) *Encyclopedia of Cognitive Science* Nature Publishing Group, Macmillan, London. Pp.

3. Szathmáry, E (2004) From biological analysis to synthetic biology. *Curr. Biol.* **14**, R145-146.

4. Lenton, T.M., Schellnhuber, H. J. & Szathmáry, E. (2004) Climbing the co-evolution ladder. *Nature* **431**, 913.
5. Szathmáry, E., Szatmáry, Z., Ittész, P., Gergő, O., Zachár, I., Huszár, F., Fedor, A., Varga, M. & Számadó, S. (2006) *In silico* evolutionary developmental neurobiology and the origin of natural language. In: (Lyon, C., Nehaniv, C. L. & Cangelosi, A., eds) *Emergence of Communication and Language*. Springer, Artington. In press.
6. Számadó, S. & Szathmáry, E. (2006) Competing selective scenarios for the emergence of natural language. *Trends Ecol. Evol.* in press.
7. Szathmáry, E (2006) Evolution. Darwin for all seasons. *Science* **313**, 306-307.
8. Y Fregnac, M Blatow, J-P Changeux, J DeFelipe, A Lansner, W Maass, DA McCormick, CM Michel, H Monyer, E Szathmary, and R Yuste (2006) Group Report: Neocortical Microcircuits -UPs and DOWNs in Cortical Computation. In: *Microcircuits: The Interface Between Neurons and Global Brain Function*. (Eds: S Grillner and A M Graybiel) MIT Press: Cambridge, MA, pp. 393-435.

PI4: Cristiano Castelfranchi

Castelfranchi C., Falcone R. (in press) *Trust Theory: Structures, Processes and Dynamics*, New York (USA): Wiley & Sons.

Conte C., Castelfranchi C. (in press) The Mental Path of Norms, *Ratio Juris*, 19 (4).

Castelfranchi C. (2006) *From conversation to interaction via behavioral communication*. In S. Bagnara & G. Crampton Smith (Eds) "Theories and Practice in Interaction Design", New Jersey (USA): Erlbaum, pp 157-179.

Castelfranchi. C., Giardini, F. Marzo, F. (2006) Relationships between rational decisions, human motives, and emotions, *Mind and Society*, 5(2).

Tummolini L., Castelfranchi C. (2006). The cognitive and behavioural mediation of institutions: Towards an account of institutional actions, *Cognitive Systems Research*, 7(2-3): 307-323.

Castelfranchi, C. (2003) *For a cognitive program. Explicit mental representations for Homo Oeconomicus*. In N. Dimitri, M. Basili & I. Gilboa (Eds.) "Cognitive Processes and Economic Behavior", London: Routledge, pp. 168-208.

Castelfranchi, C., Giardini, F., Lorini, E., Tummolini, L. (2003) The Prescriptive Destiny of Predictive Attitudes: From Expectations to Norms via Conventions, *Proceedings of the Cognitive Science Conference, CogSci2003*, Boston (USA).

Castelfranchi C. (2001) The Theory of Social Functions. Challenges for Multi-Agent-Based Social Simulation and Multi-Agent Learning. *Cognitive Systems Research*, 2(1):5-38.

Castelfranchi, C. (2000) Through the Agents' Minds: Cognitive Mediators of Social Action, *Mind & Society*, 1, pp. 109-140.

Miceli, M., Castelfranchi, C. (2000) *The role of evaluation in cognition and social interaction*. In Kerstin Dautenhahn (Ed.) "Human cognition and agent technology", Amsterdam: Benjamins, pp. 225-61.

PI5: Michel Kerszberg

Software

M Kerszberg, the *ctrl-Dev* program, an agent based application for molecular, genetic and mechanical simulations of tissues and embryonic development (<http://www.snv.jussieu.fr/~wmkersz>).

Selected publications

Kerszberg M & Changeux J-P (1998) A simple molecular model of neurulation. *Bioessays* 20, 758-770.

Kerszberg M (2003) Genes, neurons and codes: Remarks on biological communication. *Bioessays* 25:699-708

Kerszberg M (2004) Noise, Delays, Robustness, Canalization and All That. *Current Opinion in Genetics and Development* 14:440-45.

Gisiger T, **Kerszberg M** & Changeux JP (2005) Acquisition and Performance of Delayed-Response Tasks : A Neural Network Model. (2005) *Cerebral Cortex* 15:489-506.

Gisiger T & **Kerszberg M** (2006) A Model for Integrating Elementary Neural Functions into Delayed-Response Behavior. , *Public Library of Science (Comput Biol.)* 2: e25.

PI6: Arcadi Navarro

Marquès-Bonet, T., O. Lao, R. Goertsches, M. Comabella, X. Montalban, A. Navarro, 2005. Association Cluster Detector: a tool for heuristic detection of association clusters in whole-genome scans. *Bioinformatics* 21(s2):180-181.

Navarro, A., E. Gazave, 2005. Inversions with classical style and trendy lines. *Nature Genetics* 37: 115-116.

Marquès-Bonet, T. , M. Caceres, J. Bertranpetit, T. M. Preuss, J. W. Thomas and A. Navarro, 2004. Chromosomal rearrangements and the genomic distribution of gene-expression divergence between humans and chimpanzees. *Trends in Genetics* 20:524-529.

Navarro, A. and N. H. Barton, 2003. Chromosomal speciation and molecular divergence. Accelerated evolution of genes in rearranged chromosomes. *Science* 300: 321-324.

Navarro, A. and N. H. Barton, 2002. The effects of multilocus balancing selection on neutral variability. *Genetics* 161: 849-863.

AP1: Ernst Fehr

Egalitarian Motive and Altruistic Punishment, *NATURE* 433, E1-E2. (with S. Gächter).

Don't lose your Reputation, *NATURE* 432, 449-450.

The Neural Basis of Altruistic Punishment, *SCIENCE* 305, 1254-1258, 2004. (with D. DeQuervain, U. Fischbacher, V. Treyer, M. Schellhammer, A. Buck).

Social Norms and Human Cooperation, *Trends in Cognitive Sciences* 8 (2004), 185-190. (with Urs Fischbacher).

Third Party Punishment and Social Norms, *Evolution and Human Behavior* 25 (2004), 63-87. (with Urs Fischbacher).

The Nature of Human Altruism, *NATURE* 425, 23 October 2003, 785-791. (with Urs Fischbacher).

AP2: Alex Kacelnik

Approximately 140 peer-reviewed publications including:

1. Kacelnik A & Brito e Abreu F. (1998) Risky Choice and Weber's Law. *J. Theoretical Biology* 194, 289-298
2. Kacelnik A & Krebs J R (1997) Yanomamö dreams and startling prey loads: The logic of optimality. Chapter 2, In: *Human Nature: a critical reader* (Ed. Laura Betzig), pp 21-35. Oxford University Press.
3. Kacelnik A (1997) Normative and descriptive models of decision making: time discounting and risk sensitivity. In: *Characterizing Human Psychological Adaptations*. (Eds. G R Bock and G Cardew), Ciba Foundation Symposium 208, pp 51-70. Wiley, Chichester
4. Kacelnik, A. The evolution of patience. (2003). In: *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*; George Loewenstein, Daniel Read and Roy Baumeister, editors. Pp 115-138. Russell Sage publication, New York.
5. Kacelnik, A. (2005) Meanings of rationality. Chapter 2 (pp. 87-106. In: *Rational Animals?* Matthew Nudds and Susan Hurley, editors, OUP, Oxford.

AP3: Ruth Mace

Gibson M. & R. Mace (2006). An energy-saving development initiative increases birth rate and childhood malnutrition in rural Ethiopia. *PLoS Medicine* 3(4):e87.

Mace R., Allal N., Sear R & A. Prentice (2006). The uptake of modern contraception in a Gambian community: the diffusion of an innovation over 25 years. In *Social Information Transmission and Human Biology*, eds J.C.K. Wells, S.S. Strickland and K. Laland. CRC Press. (in press)

Mace R. & R. Sear (2005) Are humans co-operative breeders? pp143-159 In: Voland, E. Chasiotis, A. & Schiefenhoewel, W. (eds.): *Grandmotherhood - the Evolutionary Significance of the Second Half of Female Life*. Piscataway: Rutgers University Press

Mace R, Holden C. & S. Shennan. (Eds) (2005) *The evolution of cultural diversity: a phylogenetic approach*. UCL Press: London 291pp

Gibson M. & R. Mace (2005) Helpful grandmothers in rural Ethiopia: A study of the effect of kin on child survival and growth. *Evolution and Human Behavior* 26:469-482

AP4: Robert Rowthorn

Pagano, U., and R. E. Rowthorn (1994), "Ownership, Technology and Institutional Stability." *Structural Change and Economic Dynamics*, 5(2): 221-242

Rowthorn, R.E. (1996), "Economics and Ethics – an Economist's View" in P Groenewegen, (ed.), *Economics and Ethics*, Routledge.

Pagano U. and R. E. Rowthorn, R.E. and Pagano, U. (1996), “The Competitive Selection of Democratic Firms in a World of Self-Sustaining Institutions”, in U. Pagano and R. E. Rowthorn (eds.), *Efficiency and Enterprise Democracy*, Routledge,

Rowthorn, Robert (2006), “The Evolution of Altruism between Siblings”, *Journal of Theoretical Biology*, (forthcoming).

Guzman, R. A., Rodriguez-Sickert, C. and R. Rowthorn (2006), “When in Rome Do as the Romans Do: The Co-evolution of Altruistic Punishment, Conformist Learning and Cooperation”, under revision (in response to referees’ comments) for *Evolution and Human Behavior*

Rodriguez-Sickert, C. and R. Rowthorn (2006), “Reciprocity and Moral Norms”, submitted to the *American Journal of Economics and Sociology*.

Rowthorn R. and R. Sethi (2006), “Procedural Rationality and Equilibrium Trust”, submitted to the *Economic Journal*