examining some games to make a plausible guess. In this case, assume that the reference payoffs are the payoffs the agent would get if he were to grab at his first opportunity; that is, 5 for Al and 10 for Bob. Whatever the probability of strategies 2,1 and 3,1 (Al passes and Bob grabs), this would give Al reason for negative reciprocity amounting to a shortfall of 3, and so reduce the probability of his choosing strategies 2 or 3. This is just one illustration. In general, we can indicate reciprocity by $(Y - Y_r)(Z - Z_r)$, where Y and Z are the payoffs to Al and Bob, respectively, and Y_r, Z_r are their reference payoffs. In place of Y_i^θ in the formula for the probability of strategy i, we write $Y_i^\theta + w[sigman(Y - Y_v)(Z - Z_v)] \sqrt{}(ABS((Y - Y_v)(Z - Z)))$[EQ03], where w is a non-negative weight representing the importance of reciprocity motives to the individual. Taking the exponent 2 as above and $w = 0.333$ for a single example, we have in Table 3 the probabilities for the nine possible strategy combinations. In this case, for example, we see the overall probability of Al choosing the cooperative strategy 3 is 0.35, by comparison with 0.022 in the previous case. Note that reciprocity can lead in some cases to multiple equilibria that reinforce both cooperative and noncooperative outcomes.

We see that the probabilistic equilibrium concept admits of a larger and more plausible range of outcomes in this case than Nash equilibrium does, particularly when non-self-regarding motives are introduced in a natural way. In summary, this conception of equilibrium has three major advantages in the context of Gintis's program:

1. It allows rationality to be a relative concept.
2. Although probabilistic equilibria for some games closely approximate deterministic Nash equilibria, in some other cases, including the centipede, they can be quite different and more plausible.
3. Non-self-regarding motives are easily introduced.

Extending the behavioral sciences framework: Clarification of methods, predictions, and concepts

DOI: 10.1017/S0140525X07000787

Alex Mesoudi[a] and Kevin N. Laland[b]
[a]*W. Maurice Young Centre for Applied Ethics, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada;* [b]*School of Biology, University of St. Andrews, St. Andrews, Fife KY16 9TS, Scotland.*
mesoudi@interchange. ubc.ca
http://gels.ethics.ubc.ca/Members/amesoudi **knl1@st-and.ac.uk**
http://www.st-andrews.ac.uk/~seal

Abstract: We applaud Gintis's attempt to provide an evolutionary-based framework for the behavioral sciences, and note a number of similarities with our own recent cultural evolutionary structure for the social sciences. Gintis's proposal would be further strengthened by a greater emphasis on additional methods to evolutionary game theory, clearer empirical predictions, and a broader consideration of cultural transmission.

Gintis presents a framework for the behavioral sciences that is both rooted in evolutionary theory and based around the "rational actor," or "beliefs, preferences, and constraints (BPC)," model of human decision making. We fully support many of Gintis's aims and themes – specifically, his attempt to integrate the diverse behavioral sciences around a common framework, basing that framework in evolutionary theory (both biological and cultural) within a gene-culture coevolutionary perspective, and advocacy of the use of simple mathematical models to explain complex real-world behavioral phenomena. Indeed, there are many parallels here with our own recent articles advocating integrating the social sciences around evolutionary theory (Mesoudi et al. 2004; 2006). We maintain that stronger parallels exist between biological and cultural evolution than generally recognized, and argue that the structural and theoretical bases of the social and behavioral sciences should be mutually consistent. As noted in Mesoudi et al. (2006), individual decision-making processes will have significant effects on cultural evolution; and as Gintis notes, *culture* – the body of knowledge, beliefs, attitudes, and norms that is acquired via social learning from conspecifics – will likewise have strong effects on individual decision making and behavior. However, we believe that there are opportunities to extend the specific framework proposed by Gintis, extensions that we regard as critical to the success of his venture. As it currently stands, we fear that only a minority of behavioral scientists would empathize with his stated objectives and methods.

First, although we welcome Gintis's emphasis on evolutionary game theory, and agree that there is considerable potential for further use of this powerful method throughout the human sciences, we fear that exclusively focusing on this method may be counterproductive. We see no reason for him to limit himself to just a single theoretical technique when others – such as population genetic models (Boyd & Richerson 1985; Cavalli-Sforza & Feldman 1981; Laland et al. 1995), agent-based simulations (Axelrod 1997b; Epstein & Axtell 1996; Kohler & Gumerman 2000), stochastic models (Bentley et al. 2004; Neiman 1995; Shennan & Wilkinson 2001), and phylogenetic methods (Lipo et al. 2006; Mace & Holden 2005; O'Brien & Lyman 2003) – may be more suitable in other cases. Not all problems in the human sciences can be treated as games; many are not even frequency dependent. For example, reconstructing linguistic phylogenies using cladistic methods (Holden 2002) is a valuable evolution-inspired analysis of a problem not amenable to evolutionary game theory. Although game theory might be ideally suited to certain problems in economics or political science, it is more difficult to envisage its use in, say, neuroscience. (The link Gintis makes between strategic interactions and mirror neurons [sect. 8, para. 6] is speculative at best.)

Second, Gintis's use of evolutionary reasoning at times appears somewhat empty and rhetorical, and we encourage him to elaborate on *how* he envisages researchers could use relevant biological knowledge. His proposals would benefit from the identification of the evolutionary processes responsible for the behavioral phenomena under discussion and the generation of clear predictions amenable to empirical testing, where such objectives are feasible. For example, Gintis argues that "the economist's model of rational choice behavior must be qualified by a biological appreciation that preference consistency is the result of strong evolutionary forces" (sect. 12, para. 2), without specifying exactly what these forces are. Presumably he means some form of selection; but the selection pressures need to be atypically well-established to allow us to predict the exact preferences, beliefs, values, and behaviors that people will acquire and in what contexts. If the exercise is not sufficiently constrained by scientific knowledge, there is a danger that it could degenerate into the idle speculation and weak inferences characteristic of some modern evolutionary psychology (Laland & Brown 2002). Similarly, the statement "the ease with which diverse values can be internalized depends on human nature" (sect. 12, para. 2) fails to specify exactly what is meant by "human nature," given that human behavior varies extensively over different times and in different contexts (Ehrlich 2000), and even unlearned predispositions may exhibit considerable variation within and between populations.

Third, we encourage further clarification and qualification of Gintis's use of the phrase "cultural transmission." Gintis seemingly overemphasizes conformist horizontal cultural transmission ("cultural transmission generally takes the form of *conformism*" [sect. 1.1.2]) and mislabels other transmission biases as "conformist" (describing a strategy of "imitat[ing] what appear to be the successful practices of others" [sect. 6, para. 4] as an example of "conformist transmission," which it clearly is not). We agree that humans (and other animals) are often extremely susceptible to conformity effects, as shown by classic social psychology experiments (Jacobs & Campbell 1961; Sherif 1936). Nonetheless, the extent of conformist transmission in actual societies has yet to be empirically determined in sufficient detail, particularly relative to the numerous other social learning strategies that can be employed at different times and in different contexts (Laland 2004). Ethnographic studies often conclude that cultural transmission is vertical rather than horizontal (Hewlett et al. 2002; Ohmagari & Berkes 1997) or that considerable individual idiosyncrasy exists in culturally acquired beliefs (Aunger 2004), belying any strong conformity effect.

Many of the proposed problems with the BPC model and game theory discussed in section 9 may at least partially disappear if cultural transmission is more explicitly considered. Isolated individuals may be boundedly rational and exhibit various decision-making errors and biases, which often leads them to suboptimal behavior; but in large enough groups, a simple strategy of copying successful neighbors/behaviors can allow individuals to reach global optima quickly and cheaply. Recent experiments (Mesoudi & O'Brien, submitted) demonstrate that participants readily use and benefit from a copy-successful-individuals strategy, particularly in multimodal fitness landscapes. Gintis notes (sect. 9.6, para. 3) that scientists do not exhibit biases such as the representativeness heuristic; presumably this is because scientists have acquired the right solution (or the right means of solving such problems) from others, rather than having been born without such biases.

These criticisms aside, we endorse Gintis's scheme and hope that it provokes more interaction among psychologists, anthropologists, economists, and sociologists, as well as greater use of evolutionary theory within the human sciences.

Selection of human prosocial behavior through partner choice by powerful individuals and institutions

DOI: 10.1017/S0140525X07000799

Ronald Noë

Ethologie des Primates, DEPE-IPHC, CNRS and Université Louis-Pasteur, UMR 7178, Strasbourg, 67087 CEDEX, France.

ronald.noe@c-strasbourg.fr

Abstract: Cultural group selection seems the only compelling explanation for the evolution of the uniquely human form of cooperation by large teams of unrelated individuals. Inspired by descriptions of sanctioning in mutualistic interactions between members of different species, I propose partner choice by powerful individuals or institutions as an alternative explanation for the evolution of behavior typical for "team players."

I applaud Gintis for a brave and erudite attempt to unify the diverse and often contrasting approaches of the numerous scientific disciplines that meddle with human behavior. As a case in point he presents an array of explanations for seemingly "irrational" or "fitness reducing" contributions towards public goods shared with unrelated individuals. Altruistic punishment, an essential element of "strong reciprocity," seems Gintis's favorite example of a strategy with an irrational flavor (sect. 10. para. 5).

Here and elsewhere (Bowles & Gintis 2004a; 2004b; Gintis 2000d), Gintis endorses explanations for the evolution of such behavior in humans at two levels: (1) "classical" individual selection in repeated interactions among dyads, and (2) gene-culture coevolution with an essential element of selection at the group level. I am sympathetic to this line of thought, but nevertheless propose more reflection on other forms of cooperation before deciding that phenomena or evolutionary mechanisms require unique explanations. Cooperation can be found in a breathtaking number of forms in a wide range of organisms (Bshary & Bronstein 2004; Sachs 2004). I concentrate on cooperation between members of different species ("mutualisms," in ecological jargon), because here we also find both unrelated partners and multiple players.

Gintis points at the stabilizing effect of "punishment" in cooperating human groups (sect. 10, para 6). This reminds of a mechanism called "sanctioning" (Denison 2000; Kiers et al. 2003). Typically large and long-living individuals of one species dispose of mechanisms that allow them to reject the least profitable partners among many small partners belonging to a species with a short generation time. Examples are interactions between yuccas and yucca-moths (James et al. 1994; Pellmyr & Huth 1994) and between various plants and mycorrhizal fungi (Kummel & Salant 2006) or rhizobia (Simms et al. 2006). Sanctioning is an extreme case of "partner choice," which is a potent force of selection for cooperative behavior (Noë 2001).

There is an essential difference between punishment and sanctioning, however: selection for "selfish punishment" is favored when it changes the behavior of the punished to the benefit of the punisher. The same is true for "altruistic punishment," except that the benefit is shared by all members of the punisher's group, whereas the punisher ends up with a net cost. Punishing only pays because the punished individual is likely to interact with the punisher and/or his group again. A plant sanctioning unprofitable partners will typically not interact with the rejected individuals again. By cutting its losses, the plant obtains immediate benefits, which makes sanctioning akin to getting rid of parasites. Strong selection on the sanctioned partners is a by-product of this self-regarding strategy. The punishment concept stresses effects on individual behavior over intervals ranging from seconds to lifetimes. The sanctioning literature emphasizes the selective effect on sanctioned species at evolutionary time scales. Punishment can have a selective effect, as well, albeit probably weak, through its effect on the fitness of punished individuals. Rejection of partners in favor of more profitable ones, in turn, can shape the partners' behavior in favor of the chooser just like punishment, as long as the partners are not eliminated in the process.

Can plants sanctioning insects or fungi suggest explanations for the evolution of forms of cooperation that seem uniquely human? I concentrate on the selective force of partner choice, although selective pressure due to punishment should not be completely ignored either. A strong selective force exists when single powerful individuals (e.g., chiefs, kings, warlords, priests) or institutions (e.g., councils of elders) favor group members on the basis of their prosocial behavior. Such selection may take place during the formation of hunting parties, raiding teams, warfare, and the like. Partner choice can only be evolutionarily stable when it leads, as a rule, to an increase of fitness of both chooser and chosen. For example, the initiator of the hunt should be able to increase his individual returns by choosing the right hunters, and the participants should have higher benefits than those excluded from the hunt. This mechanism can contribute to the selection of altruistic behavior relevant to the production of public goods when partners are chosen on the basis of characteristics that make them good team players rather than good hunters per se: loyalty to the team, willingness to back up failing teammates, fairness in sharing, and so on. A recent study found that such traits are still highly relevant in modern societies: "Being a *team player* is of paramount importance in the workplace,