....

COORDINATION AND THE FOUNDATIONS OF SOCIAL INTELLIGENCE

....

DON ROSS

20.1. INTRODUCTION: THE PROBLEM OF COORDINATED AGENCY

....

The institutionalized arrangement of the behavioral sciences is based on a strong distinction between the study of individuals and the study of societies. Notwithstanding the sometime popularity of reductionism and methodological individualism, the autonomy of the social sciences from psychology and of macroeconomics from microeconomics reflects general awareness of something deeper than the fact that people behave differently in groups than they do when alone or with intimates. However, it is only recently, by comparison with the vintage of the main disciplinary distinctions, that a rough consensus has evolved over what the something in question is. Summarizing drastically, the following assumptions have become widely accepted: Social behavior involves coordination, and coordination is such a complicated set of processes that wherever one cannot abstract away from it, one enters into a distinct domain of modeling problems.

The rise of game theory, and of philosophical thinking informed by it, is historically crucial to the new preeminence of this understanding. Game theory makes explicit, to an extent grasped only fitfully before its development, that outcomes of social interaction tend to be radically unlike outcomes of individual cognition. People embedded in games with one another find themselves doing things they could not understand or predict through rationalization by exclusive reference to

their own beliefs and desires—that is, through the mere assumption of the intentional stance toward themselves. Social scientists, especially economists, had long appreciated that competitive pressures interfere systematically with individual optimization. However, game theory has given precision and detail to a key point that Adam Smith broadly grasped: Except in the unusual case of zero-sum games competition is not the opposite of coordination, but is mediated by it. Businesses selling similar products in the same market typically want to coordinate on strategies for competing with one another that do not drag them into price wars. People seeking social power aim to recruit at least some useful followers from among those they succeed in outcompeting, so political success does not best consist in the incapacitation of all rivals. And so on.

The thinker who contributed more than any other to carrying the message of the centrality of coordination from the precincts of formal game theory out into the broader community of social scientists was Thomas Schelling (1960). A key insight of Schelling's was that players' knowledge of one another's strategy sets and utility functions, no matter how complete, is insufficient for efficient coordination in many games that have multiple equilibria. Players should also model one another's recursive belief structures, using whatever knowledge they happen to possess about one another's behavioral track records and exposure to data, in search of information bases for coordination of expectations. Harsanyi (1967) launched the ongoing project of formalizing such phenomena. It is sometimes complained, especially by philosophers who seek rational foundations for all choice, that the use of exogenous focal points—recursively structured networks of players' beliefs about what other players are most likely to cognitively attend to or behaviorally respond to—has resisted such formalization. This is confused if understood in one sense. Focal points are products of shared psychology and cultural convergence. There is no reason to expect that a theory of rationality will have anything to say about their provenance, though it has much to say about their exploitation.

When modeled explicitly, the reasoning processes needed to support equilibria in beliefs seem highly cognitively demanding, even in simple two-person cases. Consider, for example, the common game structure that came to be known as battle of the sexes. Preserving the gender stereotypes that inspired that name, suppose that, in a lost age before mobile telephones, a couple who have faithfully imprinted their culture's preferred models of men and women go their separate ways in the morning, agreeing to meet to see a film in the evening. Suppose that their tendencies to thoughtlessness extend into the very practical domain, so that before parting they neglect to jointly decide which movie to see. Back when there were no cell phones there were a lot of cinemas, so as the evening approaches the members of the couple have a problem: At which cinema does each expect to find the other?

If they were a new couple who didn't yet know much about one another's preferences or histories, they might both be well advised to choose a cinema that has some distinguishing property that makes it publicly salient. For example, they might go to the largest cinema in town, or the most central. Another possible principle would be to focus on a property that is specially salient to their two-person subsociety, such

as closeness to their home. These are alternative focal points. Our couple will see a film together only if they each focus on the same principle of salience.

The example so far closely recapitulates Schelling's original thought experiments. Such instances lead immediately to reflection on conventions.[1] Treatments of the standard battle of the sexes game in introductory game theory texts typically fail to note that the sexism of the classic toy case is not gratuitous. Culturally promoted stereotypes might lead the woman to expect that the man will prefer the most violent of the available film choices. The same culture might instruct him that she will favor a romantic option. This yields the familiar matrix shown in figure 20.1.[2] In this case, if the structure of the game captures everything the players know, then the Nash equilibrium strategy is for each to randomize, while attaching a higher weighting to their own preferred alternative. Depending on the cardinality (in the von Neumann-Morgenstern sense) of their preferences, they will coordinate on the same film less than 50 percent of the time. Once again, however, sexist culture might help them to do better. If men suffer less social disapproval for selfish behavior than women, then our couple might effortlessly converge on the violent option with high or even complete reliability. (Of course, such gains in efficiency should not generally be expected to be larger than the welfare losses suffered by women as a result of sexist norms; this relationship will vary with economic circumstances.)

Two general principles for the foundations of social science have been widely absorbed from considerations of stylized problems such as battle of the sexes. First, coordination is intrinsically difficult due to multiplicity of equilibria in most interactions. Second, many features of social organization evolve and stabilize because they ease this difficulty. These principles have been widely put to work at both local and general levels of explanation. On the local level, they are often applied to explain specific social phenomena. For example, Young (1998) uses them to explain the evolution of traffic rules and norms governing contracts between landlords and sharecroppers. More broadly, they have been used to explain the general phenomenon of co-evolved sociality and extreme cognitive-behavioral plasticity—that is, intelligence—in humans.

| | | Woman | |
		Violence	Romance
Man	Violence	3, 2	1, 0
	Romance	0, 1	2, 3

2 pure strategy NE: (V, V) and (R, R); many mixed strategy NE

Figure 20.1 Battle of the sexes

The two principles as stated above are broadly correct, but they are crude. First, they do not distinguish between different ways in which coordination problems might be thought to be difficult, which may in turn imply different solution mechanisms that interact with one another in human social structures and evolution. Second, they are silent on the different possible dynamic trajectories by which coordination problems and socialization might drive one another. Did socialization come first in hominid evolution, giving rise to complex coordination problems that higher intelligence then evolved to solve? Or, in light of the importance of social norms and conventions for selecting focal points in coordination games, did intelligence generate coordination problems to which increasingly complex social structures are adaptations?

Several authors (Ross 2005, 2008, 2011; Thalos and Andreou 2009) have recently drawn attention to a philosophical blind spot in the tradition that borrows decision theory and game theory to model problems in the foundations of social science. It is standard practice to hold the identities of agents playing games as fixed throughout all stages of analysis. In classical applications players are assumed to correspond to biological organisms that persist in their agency through their lifetimes. In evolutionary games, players are lineages of linked genes—approximately, species, or frequency-dependent equilibrium strategies in within-species evolutionary games that have polymorphic equilibria. The latter approach is entirely appropriate, though, as Thalos and Andreou (2009) note, it lacks a general technique for modeling ways in which intelligent individuals or coalitions of them can subvert the strategies they have genetically inherited. The classical assumption, that individual agents are coextensional with individual organisms, is simply atomistic dogma.

A recently developed theme in some methodological and philosophical literature on applications of game theory to social science has been the human disposition to engage in *team reasoning* (Hollis 1998, Bacharach 2006). People, it is suggested, can under some circumstances at least partially fuse their agency, and evaluate alternative strategies with respect to their value to "us" or "our current project" instead of to "me." This suggestion makes many economists uneasy, because the relevant framing calls upon individuals to mysteriously forget their own interests. It is true that team-reasoning models risk being untestable unless we have empirically motivated and rigorously formulated constraints on changes in players' frames while games evolve. Ross (2005, 2007, 2008) labels this "the game determination problem" and discusses constraints on its solution. A problem with Ross's examples is that they involve complex shifts in cultural identities, which may make the phenomenon of endogenous game determination seem exotic and of marginal importance, encouraging economists to set it aside as not generally worth its cost in model tractability. Thalos and Andreou (2009) bring the point down to earth by drawing attention to a basic instance hiding in plain view: bonding between animal mothers and their offspring. Of course one *can* treat a mother bear and her cub as separate players in a game, with distinct utility functions and strategy sets. Insofar as the cub selfishly exploits its mother in certain respects, one must so treat them to capture the respects

in question. But these issues truly *are* marginal for purposes of behavioral ecology and ethology. Within those settings, modeling parsimony dictates treating the mother and her cub as a single agent, whereas insistence on methodological individualism entails pointless modeling complications and risks obscuring available generalizations. If, for some modeling purposes, it is best to represent two bears as coordinated to the point of agent-fusion, why should the same device not be available where humans are concerned?

This rhetorical question, it will be pointed out, has answers. People, but perhaps not mother bears, can always *remember* that the relevance of their distinct identities lurks in the background when they are cooperating; so the game theorist modeling their interactions should remember it too, by including the necessary variables and parameters in her model. Thalos and Andreou respond to this in a way that is not quite satisfactory. Natural selection, they say, should force humans to sometimes forget themselves for the same reason it forces mother bears to do so. Of course, they recognize—indeed emphasize—that in more cognitively sophisticated organisms the mechanisms of bonding must be and are more complex than the relatively mechanical imprinting and control by oxytocin that seems to characterize many animal parent-offspring bonds. But this recognition undermines their simple resort to natural selection in the face of the hypothetical economist's objection that unless people can forget their individuality as thoroughly as mother bears do, then modelers should not forget it either.

I will not attempt to resolve this problem in the pragmatics of game theoretic modeling here. Suppose we agree only that groups of human individuals can and often do come to have very closely aligned utility functions, and strategy sets that depend on this alignment (i.e., moves they can each make in games only by acting together). This *mere* recognition assumes away questions about whether and when coordination is difficult by treating it as accomplished from the start. The problem for the couple trying to select cinemas in our previous example was not that they lacked commonality of interests or strategy sets, but that they might lack important information needed to exploit these commonalities. We can generalize this point by distinguishing between two potential problems:

1. coordination is often or sometimes difficult because agents are selfish and tempted by opportunities to free ride;
2. coordination is often or sometimes difficult because agents have limited access to information they need in order to operate effectively as teams.

In what follows, I will set (1) aside, taking Thalos and Andreou's main conclusion as read. Notwithstanding the fact that free riding regularly disrupts or blocks coordination on particular occasions, it is an observational datum that people regularly bond with one another using a variety of mechanisms, including *identifying their own agency with patterns of performance necessary to achieve team objectives.* This leaves problem (2). As we will see, it can be especially acute when efficiency involves division and specialization of labor, that is, when successful accomplishment of team objectives requires each team member to adopt a unique strategy.

I will survey recent progress on problem (2) with regard to the following question in the foundations of social science: To what extent is the evolution of social intelligence explained by appeal to the difficulty of coordination? Thalos and Andreou reject some popular approaches in the study of the foundations of social intelligence on the basis of their view that the difficulty of coordination, in sense (1), has been exaggerated due to the influence of individualistic assumptions. I believe they are right about this. However, their diagnosis of the issue is incomplete until it is complemented by consideration of problem (2).[3]

20.2. Game Theory and the Social Intelligence Hypothesis

Among the most salient correlations in population-level animal biology is that holding between noneusocial sociality and intelligence. Consider a partial list of animal types that are highly social, not eusocial, and highly intelligent: primates, canines, hyenas, toothed whales, elephants, raccoons, pigs, corvids, and parrots. It is easy to consider for comparison a *complete* list of organism types that are highly intelligent and not highly social, because this list contains no or almost no members.[4] Finally, a complete list of organism types that are highly social and not highly intelligent includes all and only the eusocial types (termites, some hymenoptera, naked mole rats). Most types on both the first and third lists are more closely related to some types not on their lists than they are to some types with whom they share their list membership. Thus the co-occurrence involves convergent evolution and is not an accident of genetic history.

This is the observational basis for the widely accepted social intelligence hypothesis, first articulated by Humphrey (1976). According to this hypothesis, non-euscial sociality and intelligence coevolve because the main selection pressure that gives rise to advanced cognitive facility in various species is the value to individuals of remembering particular social colleagues' identities, along with their varying track records of reciprocal cooperation and tendencies to engage in conflict over mates, food, status, and other resources. A later, somewhat more specific version of the proposal is the so-called Machiavellian hypothesis, which emphasizes the importance of keeping track of who is in whose collation of alliances for the sake of political success in maintaining status and rank (Byrne and Whiten 1988, Whiten and Byrne 1997). Dunbar (1998) constructs a more precise subsidiary generalization, the social brain hypothesis. According to this hypothesis, the need for intelligence to manage larger numbers of social relationships explains observed correlation between average sizes of social groups in species and order-relative encephalization quotients (EQs), which measure the extent to which actual brain sizes differ from expected brain sizes when other relevant influences are controlled for. More specifically, according to Dunbar, EQ sets upper limits on group sizes. Surveys of recent

empirical investigations based on all of the variants of social intelligence theory are provided in Emery, Clayton, and Frith (2007).

A key commitment of social intelligence theory is that sociality and intelligence *co*evolve. This is encapsulated in the relevant idea of sociality at play in the first place. Many relatively unintelligent animals cluster together in herds for protection from predators, but do not enter into complex relationships with specific individuals as defined by properties other than consanguinity relationships (Ross 2010). A group of organisms constitutes a *society*, as opposed to a herd, when (i) the expected individual fitness of each organism is partly a function of the differential fitness of the group, (ii) individuals are to some extent restricted, by factors going beyond geographical proximity or physical barriers, in their ability to switch groups, (iii) most group members are associated with some other members, aside from mates, direct ancestors, and direct descendants, in ways that influence their expected fitness, and (iv) random re-permutations of the individuals figuring in these within-group associations would produce wide variations in fitness distributions after sex and age differences are controlled for. These properties give rise to complex interdependence that implies need for coordination.

Within social intelligence theory, the Machiavellian and social brain hypotheses attach particular emphasis to properties (iii) and (iv). What cognitive challenges are specifically implied by these properties? Authors discussing the Machiavellian and social brain hypotheses often write as if the main pressure arises simply from the need to remember the identities and dispositions of many individuals. However, the extent to which this taxes neural functioning is not obvious, and cannot validly be inferred from the large experimental literature showing that people have difficulty holding significant numbers of unrelated objects in simultaneous conscious attention. Mammal brains have prodigious capacity for unconsciously remembering lexical associations, thanks to parallel distributed processing and massive neuronal and synaptic capacity (Deacon 1997). For example, squirrels, which are not especially intelligent mammals, store impressive lists of food cache sites in memory (though not without errors).

From the game-theoretic perspective, we can identify two approaches to cognitively representing coordination dynamics in sets of interacting individuals with distinctive strategic dispositions. One approach simply stores and updates lists of condition-action pairs. Thus, for example, an animal might tag each conspecific with which it has a strategic relationship with labels such as "reciprocated cooperation last time" or "failed to reciprocate cooperation last time." Updating of such tags on each episode of interaction could readily facilitate use of simple but relatively robust strategies such as tit for tat in coordination games with competitive aspects. Again, however, it is not clear that storing and using lists requires unusual intelligence.

The alternative approach to cognitively controlling one's own interactive dynamics with specific individuals is to compute representations of, and solutions to, extensive-form games. This might be done using heuristics that are less reliably accurate and powerful than a game theorist's trees or sets of equations, but on any implementation it is likely to require considerable neural processing resources.

However, for the analyst, fleshing out the Machiavellian or social brain hypotheses using this cognitive conjecture faces two serious empirical problems. First, observed instances of so-called reciprocal altruism among nonhuman animals are largely limited to interactions most naturally characterized using binary strategy sets (e.g., regurgitate blood for another vampire bat or don't; guard another's offspring or ignore them) and repeated iterations of identical stage games. It is unclear in such cases why representation of extensive-form games would have any advantages over, or be behaviorally distinguishable from, simple condition-action list storage. Second, numerous experiments show that *people* do not, except after substantial situational training (Kagel and Levin 1999; List and Lucking-Reiley 2002; Levitt, List, and Sadoff 2011), choose strategies using the kind of backward-induction logic that game theorists rely upon for solving extensive-form games. A particularly persuasive demonstration of this, using MOUSELAB, is given by Camerer et al. (1993). And it is unlikely that any nonhuman animals are better *theoretical* strategists than humans.

Thus *one kind* of coordination that seems to be difficult for untrained humans, and by implication for animals in general, is coordination on *socially efficient* equilibria *in pure strategies* while playing *one-shot* games with *new partners*. This should not be assumed to simply be the limiting case of a more general fact that all coordination is difficult. A background contributor to the difficulty that untutored natural agents encounter in finding subgame perfect and sequential equilibria[5] may be precisely that they usually find it *easy* to locate many equilibria involving mixed strategies, especially where these make allowance for quantal responses[6] (McKelvey and Palfrey 1995) by statistical tuning. This can lead groups of players to be captured by such equilibria. Game theorists, by contrast, identify the hard-to-find pure-strategy equilibria by deduction.

How might natural agents find it easy to implement mixed-strategy equilibrium play? A research program initiated by Paul Glimcher and collaborators (see Glimcher 2003a), based on single-cell recordings in monkeys, strongly suggests that, at least in primate brains, individual neurons in the circuit that estimates comparative reward values directly compute statistical variations in choice that track Nash equilibrium (NE) mixtures.

In Glimcher's basic paradigm, monkeys are trained to implement choices by directing their gaze to one member of a set of colored flashes on a computer screen. While the monkeys do this, activity in single neurons in the lateral intraparietal area (area LIP) is recorded. Neurons in this area encode salience of visual targets, and thereby direct attention to them. They deliver output to parts of the visuomotor system that plan and execute eye saccades. Thus it is hypothesized that area LIP neurons do not compute, but are closely correlated with, the monkeys' decisions about where to look. One of Glimcher's early breakthrough experiments (Glimcher 2003b) found that when monkeys played a game involving competitive coordination against a computer (the so-called inspection game shown in figure 20.2), firing rates of LIP neurons were equal for each pure strategy that was mixed in NE. If the neurons were tracking probabilities of movement instead of expected utility, this

should not have been observed. This interpretation of the observations was greatly strengthened and enlarged in importance by a subsequent experiment (Dorris and Glimcher 2004) in which monkeys learned changing reward values in the same game. Trial-by-trial fluctuations in LIP activity correlated with trial-by-trial behavioral estimates of expected utility.

Work by Lee et al. (2004) has extended the implications of this result where coordination in games is concerned. Lee's group investigated the performance of non-human primates in the matching pennies game (figure 20.3) played against a computer. In the figure, H and T simply denote heads and tails in a coin-flipping task.

In this game, one player is incentivized to coordinate and the other player is symmetrically incentivized to avoid coordination. Of course, the anticoordinator

NE = (randomize, randomize)

Figure 20.2 The inspection game

NE = (randomize, randomize)

Figure 20.3 Matching pennies

must identify coordination, at least implicitly, in order to avoid the other player's effort to achieve it. The game's unique NE is mutual randomization. This may make it appear unpromising as a test-bed for an experiment on equilibrium selection. However, by locking the computer into non-NE strategies, or depriving the computer of the ability to punish non-NE play by its opponent, Lee's team could present monkeys with alternative best-reply strategies.

Monkeys were divided into three groups, each of which played a computer implementing a different algorithmic strategy. Algorithm 0 played the mixed NE strategy unilaterally, without looking for patterns to be exploited in the monkeys' responses. In such cases monkeys show strong biases toward one or another of the pure strategies. (The source of these biases in the experiments is unknown and was not investigated.) Algorithm 1 detects such biases and counteracts them by driving the computer toward the opposite bias. Monkeys respond by dynamically conditioning switches between pure strategies on specific wins and losses. This win-stay, lose-switch (WSLS) strategy is an implementation of matching behavior, which has long been identified by animal learning theorists as the default strategy adapted by most learners confronted with uncertainty in alternative reward streams (Herrnstein 1961, 1982). It is as good a strategy as any other against Algorithm 1. Finally, Algorithm 2 looks for patterns in recent histories of monkey play compared with the monkey's gains and losses and exploits any such patterns. The only NE response to Algorithm 2 is randomization between pure strategies. Monkeys learn this.

A follow-up experiment (Lee et al. 2005) applied the same protocol to the familiar rock-paper-scissors (R-P-S) game, which exactly reverses the logic of coordination. In response to Algorithm 0, monkeys again tended to settle on a preferred pure strategy. In response to Algorithm 1, they played the so-called Cournot best response, biasing their choices in each round n by reference to the pure strategy that would have won or did win in round $n-1$. In response to Algorithm 2, monkeys approximated but didn't quite achieve randomization. In particular, they came as close to randomization as implementation of a classical Rescorla-Wagner conditioning rule can get. This is highly suggestive and should mitigate initial surprise at the monkeys' demonstrated skill with statistical induction. Game strategy choice that is not based on backward induction or conjectures respecting Bayes' rule can very closely mimic optimal play simply through standard conditioning. Outside of some rigorously monitored asset markets and game theorists' experimental labs, it is unlikely that humans, let alone monkeys, often face opponents that can detect and exploit the difference between true NE play and NE-mimicry.

These results motivate search for a neural mechanism that will adjust stochastic response frequencies toward WSLS when the environment strategically adapts to it but doesn't model it, and implements reinforcement learning when the environment strategically models it. Direct recording has identified individual neurons in dorsolateral prefrontal cortex (dlPFC) that modulate their activity in ways that allow for trial-by-trial comparison of two alternatives, as in matching pennies (Barraclough, Conroy, and Lee 2004; Seo, Barraclough, and Lee 2007). Crucially,

some dlPFC neurons store "eligibility traces," that is, modulate their response probabilities in light of results of previous responses. This would be necessary in a neural system able to learn best (or near-best) replies to Algorithm 2.

Lee and Wang (2009) review models of processes by which some individual DLPFC neurons might adjust their stochastic response probabilities in such a way as to implement WSLS when randomization doesn't improve on it, and randomize when the environment responds as if strategically modeling the organism's behavior. The first is a specific type of "ramping-to-threshold" model based on drift diffusion. Suppose we have two alternatives X_1 and X_2 for comparison, where $X = X_1–X_2$. Then the dynamics of X are modeled by drift diffusion if

$$dX/dt = \mu + \omega(t)$$

where μ is the drift rate and $\omega(t)$ is a white noise of zero mean and standard deviation σ. μ represents a bias in favor of one alternative X_1 or X_2. The system is a perfect integrator of the input

$$X(t) = \mu t + {}^t \int \omega(t')dt'$$

and terminates whenever X(t) reaches a positive threshold θ (choice 1) or $-\theta$ (choice 2). If μ is positive then choice 1 is correct, while choice 2 is an error; otherwise the opposite. If μ is 0 the system is "set" to randomize.

Can neural circuits implement this model? Lee and Wang point out that they can't perfectly integrate inputs, because they "leak"—that is, drift with time independently of μ. (That is to say: They forget.) However, Wang (2001, 2002) and others have demonstrated that this can be corrected by recurrent activation. So as long as input meets some persistence threshold, and the neuron is embedded in a network that receives stabilizing feedback from elsewhere in the brain (i.e., there is recurrence), then a neuron's stochastic response rate can systematically adjust in the way approximately described by the drift diffusion model. If neurons ramp to threshold, then learning must involve adjustments to the thresholds (e.g., to θ in the model). Lee and Wang suggest that this is what the dopamine signal in the midbrain reward circuit does.

We can shift to a less abstract level of modeling—but constrained by the higher-level models—in search of greater biophysical accuracy. A *spiking network* model takes account of the specific rise-times and decay-times of synapses. Lee and Wang point out that "synaptic dynamics turn out to be a crucial factor in determining the integration time of a neural circuit dedicated to decision making, as well as controlling the stability of a strongly recurrent network" (2009, 499).

In general, what hold stochastic neural firing frequencies within stable but adjustable bands are presumed to be network properties that create complex dynamics with multiple attractors. Pairs of neural networks linked to one another by both excitatory and inhibitory connections decide between two alternatives A and B, with C_A and C_B denoting recurrence of input favoring A and B respectively, according to the softmax function

$$P_A(C_A - C_B) = 1/(1 + \exp(-(C_A - C_B)))/\sigma$$

where σ denotes the "extent of" stochasticity in the network. Such a system will decide between A and B even when absolute magnitudes of C_A and C_B are small, and when $C_A = C_B$. Lee and Wang claim that the function performs well in describing monkey behavior in the matching pennies game, but hasn't yet been tested on single neuron responses in DLPFC during game play. They report that it has produced good fits in estimating firing rates in area LIP when monkeys learned to anticipate stochastic changes in visual displays. However, it should be noted that this claim is too imprecise as it stands to be empirically evaluated. Lee and Wang do not discuss possible ways of generalizing the softmax function to describe dynamic attraction that would incorporate choices among more than two alternatives, as in their R-P-S experiment. Nor do they mention any test of a relationship between their model and the reported pattern of monkey play in R-P-S.

Lee and Wang introduce an additional methodological strategy by using their model to generate a simulation of neural learning of the monkeys' task in the matching pennies experiment. It captures the broad characteristics of the observed behavior, though it underpredicts the monkeys' frequency of use of WSLS against Algorithm 1. To reproduce this, Lee and Wang adjust the model by using "a different learning rule, according to which synapses onto both neural populations (selective for the chosen and unchosen targets) are modified in each trial. This is akin to a 'belief-dependent learning rule'" (2009, 498). They report that the model used for the simulation required different parameters for play against each of the three algorithms, and that they "incorporated a meta-learning rule proposed by Schweighofer and Doya (2003) that maximizes long-term rewards" (Lee and Wang 2009, 498). Thus the simulation really tests a *joint* hypothesis: that some neurons in the reward system tune their response probabilities by ramping to thresholds via strongly recurrent dynamics, thereby learning strategically superior stochastic behaviors, *and* parameters that govern applications of this learning model are themselves learned by some so far unspecified neural process somewhere else in the brain that implements the Schweighofer-Doya meta-learning rule.

The evidence to this point thus suggests that individual neurons in the primate reward system, possibly supplemented by more distributed encoding of some learned parameters, steer the organism as a whole to approximate NE behavior in basic coordination games. This hypothesis incorporates a significant element of speculation, due to its partial reliance on simulation. And one naturally wonders whether significant differences would be observed if the experiments with monkeys were replicated in less social or less intelligent animals. On the other hand, the part of the hypothesis that is on firmest empirical ground has portentous implications for social intelligence theory. This is that solutions, or at least behaviorally similar near-solutions, to basic coordination problems are computible as statistical exercises by primate brains and do not need to rely on culturally evolved focal points. Furthermore, the kind of coordination that even people find very difficult, collectively identifying equilibria by backward induction in novel, extensive-form games

that depend on modeling specific, idiosyncratic partners, seems *too* difficult to have pushed hominids up the ramp to highly social intelligence, because even modern people's untutored and unscaffolded cognitive dispositions are not up to the job.

Thalos and Andreou (2009) are skeptical about social intelligence hypotheses on the grounds that such hypotheses begin from the assumption that coordination arises when utility-maximizing *individuals* are confronted with political challenges. They maintain instead that the foundation of the extreme social integration found in humans lies in the fact that the evolution of coordination describes processes by which people evolved to face their most complicated social situations not as individuals but as members of teams. They therefore reject accounts that depend too heavily on internal cognitive representation of coordination dynamics by individuals. Human coordination plausibly arises, they argue, from emergent dynamics without need for sophisticated cognitive representation; to this extent, humans more closely resemble eusocial animals than most of the literature supposes. One obvious objection to this thesis is that whereas we know roughly how ants and bees chemically transmit the information to one another that is necessary to support the coordination they noncognitively achieve, it is implausible that humans could informationally support their much more complex coordination dynamics using only low resolution subcognitive signaling such as odors and posture.[7] The findings in monkeys as reviewed above go some way toward answering this objection. All primate—and, for that, mammalian—midbrain reward systems are anatomically alike, and most evidence to date at least weakly supports functional similarity. Thus the single-cell work with monkeys is evidence that people's brains contain neuronal groups that are, in effect, automatic mixed-strategy equilibrium calculators. In one respect this motivates moderating Thalos and Andreou's hypothesis. If human brains calculate NE of games amongst individuals, then this implies that in some sense brains do represent their social situations in terms of such games after all. On the other hand, the representation in question could be—in the case of the monkeys, surely is—subcognitive and distributed. Since the reward system appears to be encapsulated from conscious representation and deliberation, allowing that it solves games among individuals as input to further frontal processing, where reframing in social context might occur, seems to be consistent with Thalos and Andreou's intended main emphasis. This concession would begin to address the serious problem with their hypothesis that its informational basis is completely unspecified.

This does not see off the whole of the skeptic's objection, however. We noted earlier that an economist might defend a strongly cognitivist model of game representation and solution in a way that resembles the standard defense of rational expectations in macroeconomics. Ants and bees probably *cannot* represent themselves as independent agents, whereas humans evidently can and do so conceive of themselves. Therefore, we should be at least uneasy with the suggestion that they forget to so conceive of themselves precisely in the sorts of situations, social exchanges, where members of our species typically invest their largest personal stakes. This is the sort of objection, however, to which the evidence about people's failures to use extensive-form representation and the backward induction it

supports is precisely relevant. To the extent that coordination dynamics among people rely on reframing of agency buttressed by subcognitive prediction of outcomes based on conditioned Rescorla-Wagner learning and drift diffusion, we should predict that people would find everyday coordination in relatively structurally simple games easier than defenders of social intelligence hypotheses generally assume, but that efficient coordination in games with relatively more complex informational dynamics might often or usually fail to be achieved. We will consider further evidence for this in the next section.

A further respect in which the neuroscientific experiments with monkey strategic choice may be taken to support Thalos and Andreou's thesis is that the mechanism, in representing and solving games as statistical rather than logical problems, is more naturally adapted to coordinating the individual's behavior with that of a group, rather than to supporting idiosyncratic models of specific individuals. This theme will also loom large in the next section of the chapter.

Merlin Donald's (1991) theory of the evolution of the modern human mind, based on interpretation of physical and cultural anthropological evidence, can be retrospectively interpreted as providing further support to Thalos and Andreou's thesis. Donald argues that the career of *Homo erectus* involved a major transition from the *episodic* representations of social situations upon which contemporary great apes rely to *mimetically* structured representations. The latter are presented as a necessary platform for the later transition to the fully semiotic, abstract representations characteristic of the modern human mind and expressed in modern humans' distinctive use of languages with structured grammars. The core difference between an episodic and a mimetic representation is that the latter, but not the former, involves perceiving and storing memories of specific behaviors of others by reference to general stylistic features that allow them to be subsequently reenacted. This in turn provides a basis for limited cross-generational learning and cultural accumulation. Donald argues that this significantly increased the selection advantage of greater memory capacity, and that the flowering of mimesis in *H. erectus* thus predicts and partially explains the major advance in encephalization that make this species the pivotal anatomical transition figure between apelike hominids and modern humans. We cannot here devote space to reviewing the rich mix of evidence that Donald's theory integrates. The important point for present purposes is that, according to his well-received account, imitation, though it rested on prior increases in cognitive capacity in the hominid line, was a necessary precursor to the runaway evolution of intelligence that radically distinguishes late *H. erectus* and his descendents from all other primates. Imitation, especially of the fully mimetic variety that does not require the immediate presence of the imitated party's behavior, is the simplest and most direct possible vehicle for coordination.

Donald of course does not imagine that mimesis could have evolved in a species that was not social, or not highly socially intelligent. His account of the sociality-intelligence dynamic is therefore, like almost all competing stories currently taken seriously, a coevolutionary one. However, we can say that it presents sophisticated coordination as an enabler of (rather than a consequence of) distinctively *human*

intelligence, since according to Donald it was a qualitatively novel form of coordination that provided the platform for a qualitatively novel level of niche transformation, opportunity conception, and problem solving, rather than the other way around. It is for this reason that I interpret his theory and the evidence he assembles for it as supportive of Thalos and Andreou's main conjecture.

However, introducing Donald's theory into the picture directs attention to complicating considerations. Thalos and Andreou cite with approval Haim Ofek's (2001) account of the role of specialization of labor and of markets in the evolution of *both* human sociality and human intelligence. Unfortunately, instead of acknowledging complexities that Ofek's thesis raises for theirs, they attempt to brusquely expropriate it by first raising a specious objection to details of Ofek's historical conjecture,[8] and then answering their own objection by simply shoehorning Ofek's thesis into theirs despite the fact that his text provides no support for their doing so.

Ofek's main contribution is to offer evidence for introducing market exchange as a third element in the coevolutionary matrix that produced *H. sapiens*'s remarkable ecology. Reasoning carefully from the range of economic problems that faced prehuman and early human hunter-gatherers, Ofek builds a case that development of the metabolically expensive brain required dietary shifts that could only be supported by the development of specialization of labor and exchange across bands. The earliest tradables produced by specialist producers, Ofek suggests, were fire-maintenance services[9] and sheep and goat herding. Cross-band exchange in turn required the partial displacement of natural xenophobic violence by diplomacy, thus promoting social intelligence and the enhanced strategic competence in which it partly consists.

Specialization and market participation are forms of coordination that cannot be based on imitation alone, since they crucially depend on polymorphic strategy selection in the population. At the very least, they also require equilibrium learning; families should have imitated successful fire-keepers only up to the point where the opening of one more competing fire service station equalized the expected utility of the next fire-keeping family and the expected utility of remaining hunter-gatherers. Now, it has already been suggested that Thalos and Andreou's thesis is strengthened, although in a more complicated version, if we take account of the evidence for sub-cognitive computation of equilibria by primate neuronal groups. However, the Glimcher and Lee and Wang experiments test only for computation of equilibria in games involving two players where information is perfect and all moves are directly observed. *Homo erectus* families might have found labor market equilibrium much harder to track. And this was the very dawn of labor market formation; as Seabright (2010) discusses, the complexity of such problems swiftly exploded for people, and still continues to accelerate.

The point here is that the evolution of human coordination capacities was not simply a single ascent up one complexity gradient. Social intelligence hypotheses are intended as accounts of the early coevolution of sociality and intelligence that facilitated team reasoning in small family bands. In my view, the considerations surveyed in the present section provide persuasive motivation for suspecting that

Machievellian arms races were at best a small part of this story. Coordination, including coordinated competition, was simply not as difficult for our ancestors, even our relatively small-brained ones, as tellers of social intelligence tales often suggest. However, to confine attention to the period before small bands of hunter-gatherers began to coalesce into large, settled societies is to avert our gaze from the most dramatic and biologically revolutionary episodes in the distinctively *human* career of socialization. The specialized division of labor became a globally transformative force once it was applied beyond simple barter exchange between kin groups. The basic dynamic of coordination, and its relationship with socialization, changed equally profoundly. The logical basis of this has already been identified above: As market pressures and opportunities begin to dominate social organization, people must abandon coordination based on simple imitation, lest everyone be trapped in buying high and selling low. In the concluding section of the chapter, I will suggest reasons for thinking that the more complicated demands of market coordination provided the impetus for a novel twist in the evolutionary dynamics of coordination capacities, one that turned in the opposite direction from that emphasized by Thalos and Andreou: People learned to construct themselves as distinctive individuals who could participate in multiple teams at the same time.

20.3. GLOBAL GAMES, OVERCOORDINATION AND THE VALUE OF INDIVIDUALITY

The previous section surveyed responses to a bias in intellectual history, mainly derived from the extension of game theoretic reasoning into the foundations of social science. The bias in question exaggerates the difficulty of equilibrium selection in normal-form games of complete information among small numbers of players.[10] The reader might reasonably feel some surprise at this. Since when did economists, of all people, get stuck in cul-de-sacs by *underestimating* human capacities for tracking statistical regularities? Economists, after all, are the people who have given us such virtuosos of instantaneous complex computation as participants in Walrasian auctions, holders of rational expectations, and infinitely lived participants in efficient markets.

The answer, already hinted at, is that economists and game theorists didn't underestimate pre and early humans' computational abilities. (Few have directed their attention to these agents at all.) Rather, they encouraged relatively thoughtless projection, by other behavioral scientists, of the more complicated coordination problems that preoccupy current policy engineers to the early stages of hominid social formation. In the kinds of coordination games that most interest economists, agents' choices, at least collectively, exert more power over available outcome spaces than do the choices of monkeys playing matching pennies against preprogrammed opponents. This power destabilizes agency itself, by making utility functions

dynamic and by embedding games within meta-games. The evolution of modern societies is characterized by dizzying acceleration in the special human capacity for niche construction; by their behavior traders don't merely adapt to markets, but change their structures. This raises the following problem for the game modeler. If, as we should expect to be typical, people approach their strategic interactions with both asymmetric information and the ability to exploit this information to dynamically influence outcome spaces, why and how should we imagine that agents converge on a shared model of outcomes? Such convergence is a necessary condition on our being able to say anything about players' levels of knowledge of the structure of their games.

Morris and Shin (2003) note a particular path by which economists have introduced indeterminacy in outcome identification into many—indeed, most—models of coordination around investment and saving behavior. We begin by supposing that all agents deduce correct beliefs about the distributions of beliefs, and of beliefs about beliefs, and so on, among all agents involved in a game. Now imagine that, as is common in market interactions, some beliefs are self-fulfilling: Agents choose actions on the basis of their beliefs about beliefs, others correctly believe they will take the actions in question (because their belief assignments are also correct), and then the actions in question, when taken, confirm the beliefs. Finally, suppose that all exogenous aspects of the state of the economy are common knowledge. This set of standard assumptions, made partly for the sake of tractability, but also because alternative assumptions are arbitrary in the absence of special considerations, produces perfectly coordinated beliefs and actions. In that case we have no means of discriminating different probabilities associated with different equilibria unless we know how the agents arrived at their initial beliefs before they computed the solution to the game. These beliefs drive everything, and so do *all* of the work in selecting an equilibrium. But of course we generally do not, and cannot, have empirical access to processes by which agents arrived at their initial beliefs.

This problem doesn't rest on the assumption that coordination per se is unrealistically difficult. Rather, it makes outcome prediction based on the assumption that coordination will be achieved unrealistically difficult.

A technology that tries to break out of this kind of cul-de-sac, developed by Carlsson and van Damme (1993), is known as global game theory. In a global game, players receive slightly noisy, nonpublic, signals about uncertain states of the economy. If players have correct beliefs about the sources of noise, when each one observes his own signal he can estimate the distributions of signal values received by other players. Not knowing their background beliefs, he assumes that these are randomly distributed about the unit interval, because in his ignorance this is the least arbitrary prior. (Morris and Shin say that it resembles Laplace's principle of insufficient reason: When no consideration favors one possibility, assume equiprobability.) On this basis the player estimates the probable distribution of actions by others and chooses his best reply. Carlsson and van Damme show that given some plausible technical restrictions, this setup mimics the solution space of standard, pre-epistemic game theory, while nevertheless taking into account that players

choose actions in light of their beliefs about the beliefs of others. Put less impressionistically, and following Morris and Shin's (2003, 86) exposition, suppose we have the game shown as figure 20.4. Each player I observes a signal $x_i = \theta + \sigma \epsilon_i$, where the ϵ_i are eight-dimensional noise terms. So σ parameterizes an incomplete information game. If the payoff vector $\theta \in R^8$ is drawn according to a strictly positive, continuously differentiable, bounded density on R^8, and the noise terms for each player are drawn according to a continuous density with bounded support, independently of θ, then as $\sigma \to 0$ any sequence of strategy profiles that survives iterated elimination of strictly dominated strategies converges to a unique limit, independent of the noise distribution, which includes the unique NE of the underlying complete information game if it exists, and the risk-dominant NE if there is more than one strict NE.

A major application of global game theory has been to speculative crises in financial markets, which in the standard game-theoretic models are unpredictable because of multiple equilibria. Morris and Shin (1998) develop a global game in which a bank run becomes a unique equilibrium if a parameter driving expected rate of return for longer-term depositors falls below a specifiable threshold. Critics (beginning with Atkeson 2001) have raised doubts that global games are actually up to the task of predicting banking crises in real markets, but we may pass over this problem when our interest, as here, is merely in qualitative representation of informational dynamics. The key points for present purposes are as follows. First, global game theory provides a formal framework for representing the way in which agents with imperfect information can converge on a common model of their games. To this extent it mitigates the apparent difficulties raised for coordination by information asymmetries. However, global game theory is also useful in helping us to describe the basis for a different problem that may beset populations that are driven to coordinate by group selection dynamics (i.e., property [i] among our previously identified conditions for sociality) and which are therefore also under pressure to coordinate on relatively *efficient* equilibria. This is that global game play, by eliminating indeterminacy, helps to show how coordinators can converge on inferior equilibria, or traps—for example, bank runs that can only be stopped by exogenous interventions.

The difficulty associated with coordination can thus be reframed. Much of the social intelligence literature imagines agents with given, relatively selfish utility

	1	0
1	θ_1, θ_2	θ_3, θ_4
0	θ_5, θ_6	θ_7, θ_8

Figure 20.4 Incomplete information game with payoff vector $\theta \in R^8$

functions struggling to compute coordinated strategies. By contrast, our attention is now focused on agents who find coordination relatively straightforward, but who partly for that reason may overcoordinate—that is, converge on an inefficient game structure and eliminate strategic variance within the population that might otherwise have provided the basis for discovery of paths to sets of equilibria containing superior outcomes.

Note that this problem is related but not identical to the less abstract herding problem that has been studied by an extensive experimental literature descended from a classic 1992 paper of Banerjee. The herding problem is that agents who are unsure about the distribution of private information in a market can, under certain plausible conditions, rationally choose to imitate strategies of other participants when they have misidentified others' herding behavior as revealing private information, and thereby be led to ignore their own genuine private information. This implies that their information is lost to the market's available information set, which implies inefficiency and can amplify through its effects on other participants' actions, creating cascades of inefficient information management. Furthermore, if initial observers happen to be unlucky, the result can be a "reverse cascade" in which everyone converges on an incorrect model.

Experimental tests of herding have generated equivocal but interesting results. Anderson and Holt (1997) found significant cascading and reverse cascading even when participants knew that they were at no disadvantage with respect to the quality of their private signals. Sgroi (2003) replicated this result with endogenous timing, that is, where subjects could decide to wait to choose until they had observed choices of others. Sgroi also tested the effect of correcting errors incorporated in reverse cascades. In these instances, participants tended to move further away from Bayesian rationality, suggesting failure to fully recognize that rational choice can produce suboptimal outcomes. On the other hand, Huck and Oechssler (2000), Noeth et al. (1999), and Spiwoks, Bizer, and Hein (2008) found general failures of Bayesian rationality, overweighting of private signals, and therefore fewer cascades.

Two experimental reports are especially interesting in the present context. Hung and Plott (2001) found prevailing near-rationality, and broad confirmation of Anderson and Holt's findings, when subjects were encouraged to frame their decisions as individuals. (Near-rationality refers to the fact that subjects produced fewer cascades than fully rational agents would be predicted to do.) However, when they understood that the majority decision would bind all participants, and were thus incentivized to reframe the choice problem as one confronting a team, subjects paid more attention to their private signals. This contrarian behavior will tend to improve social efficiency in a very noisy environment, while lowering it in a highly transparent one. However, as a further complication, when Corazzini and Greiner (2007) encouraged subjects to frame their choices in the familiar context of independent choices over lotteries, herding collapsed and individually irrational but socially efficient contrarian behavior abounded.

The logic of the herding problem applies to market efficiency *after* agents have estimated the structure of their game through global game play. This is why it is a

more interesting problem for application to markets governed by actually institutionalized rules. The problem of inefficient convergence in global games is relevant to potential selection of inefficient institutions in the first place. Now recall, from section 20.2, the conditions that govern the evolution of sociality, where the inclusive fitness of individual agents is conditioned by correlation between strategies and genetic relatedness of interactors. Thalos and Andreou follow current fashion in using the phrase "group selection" here, though this invites many common confusions (West, El Mouden, and Gardner 2010). People who structure their environments using more efficient market institutions will tend to outcompete people who rely on nonmarket institutions or, more realistically, less efficient market rules.

This hierarchy of global and local games, based on different scales of analysis where both time and agency are concerned, allows for qualitative dynamics in which agents disposed to contrarian behavior give other agents incentive to favor them because they thereby reduce the relative size of the basin of attraction of over-convergence in the space of evolutionary global games. Put in plainer terms, where the dynamic of socialization leads to evolutionary competition among groups that have developed market exchange and specialization of labor, pressure can arise that favors the cultivation of a form of *socialized* individuality that is distinct from genetic individuality. The individuality in question is socialized because it is a set of dispositions to favor certain strategies over others in global games; thus it presupposes and arises from complex sociality.

Since contrarian behavior can lead to inefficiency as markets become more transparent, the evolution of socialized individuality does not counterfactually predict the triumph of efficient markets. Indeed, models can be written in which the market volatility caused by contrarian behavior—which Thalos and Andreou might characterize as the disintegration of teams—serves as a limiting factor on the evolutionary pressure, at a coarser scale, for wider dispersion of types of individuals at the population level, with the richer space for specialization and market complexity that such dispersion implies (Richerson and Boyd 2005).

Socialized individuals are not merely genetic individuals who can refer to themselves, as fantasized by thinkers in the venerable Western philosophical tradition of social atomism. Normal human individuals are adaptations to social pressures to conform to prevailing norms but, within those limits, to cultivate patterns of distinctive, re-recognizable (and therefore relatively predictable) self-models that govern their own behavior. Such selves can perform the unique trick, valuable only in the context of the unusual human ecological strategy, of facilitating division of labor within *novel* and *creative* group enterprises. It was likely our ancestors' capacity for such joint creativity that allowed us to survive the concertina of glacial advances and retreats during our species' early career—and, by standard Darwinian logic, those same pressures that made us the unique virtuosos of coordination that we are.

One among several prerequisites for such enterprises are signaling systems—languages—that stabilize ranges of possible signal meanings by digitalizing reference. That is, human language enables one to communicatively refer to "Napoleon"

exactly, not just to an indefinite range of things sharing to various degrees Napoleon's analog blend of properties (i.e., "napoleonishness"). Thus humans can coordinate on plans involving hypothetical objects picked out by digital contrast with other members of classes into which the grammars of public languages permit them to be sorted (Ross 2007). Some philosophers (e.g., Searle 1995), in their abiding preoccupation with reference, have been overimpressed by this, suggesting that language plus shared perceptual saliences are sufficient to account for people's ethologically unique capacity to coordinate. But this is yet another aspect of the intellectual legacy that makes *basic* coordination seem harder than people actually find it to be, imagining them to begin as isolated in interior worlds that they must struggle to render commensurable. (Consider, for example, Wittgenstein's [1953, 100] famous beetle in the box thought experiment, which draws upon a theme central to Western philosophy since Locke.) The evolution of fully symbolic representation in people was enormously important to *intergenerational* coordination that, by conserving cultural and technological discoveries, accelerated human niche construction. But this is not the sort of basic coordination on which philosophers inspired by game theory—for example, Lewis (1969) and Skyrms (1996, 2010)—have so far concentrated.

As argued by a range of theorists including Bruner (1992, 2002), Dennett (1991), Turner (1998), and Ross (2005, 2007), a crucial evolved basis for guiding socialized individuality, and consequent approximate equilibria in networks of behavioral expectations, are common perceptions of narrative coherence. People insist that others with whom they enter into coordination exercises tell dramatically structured, publicly ratifiable stories about themselves and conform their behavior to these stories. This enables not only public predictability, but, for organisms with enormous information-processing devices that are largely opaque to their direct inspection, *self*-predictability (Dennett 1991).

People mutually ease the imposed burden of consistent self-construction by assisting each other as coauthors of narratives, recording expectations, rewarding enrichments of each other's subplots, and punishing overly abrupt attempts to revise important character dispositions. Parents initially impose this regime of self-construction on their children, later handing over primary control (often with much resistance) to their offspring's peer groups. Thus people become and remain distinct while simultaneously remaining comprehensible and predictable.

The fact that self-creation and self-maintenance are *achievements* requiring effort is what explains prevailing *normative* individualism. Socialized individuals are centrally important to people partly *because* they don't just drop out of the womb. A person's own social self is of particular importance to her because if its integrity is cast into doubt her capacity to enter into mutually beneficial exchanges—the fundamental survival strategy of *Homo sapiens*—is threatened. The integrity of other people's selves is valuable to each individual because this is the basis for convergence in global game play, and then for coordinating without sacrificing specialization and all private information to imitation.

What neither theory nor experimental evidence yet support, however, are general theorems on the social welfare properties of varying levels of entrenchment of

socialized individuality. Self-construction necessarily reduces the strategy sets available in coordination dynamics, since the predictability of the socialized individual depends on foreclosing some possibilities for action that would otherwise be available. At the same time, it reduces the loss of private information in imitation cascades by facilitating adoption of contrarian behavior in the sense of the herding literature. Socially optimal trade-offs between these effects are almost certainly highly sensitive to specific levels of risk, environmental change, and transparency of information about contingent relationships between present states and possible outcomes.

20.4. CONCLUSION

Methodological individualism is currently a minority position in the philosophy of social science. The majority of economists likely still pay lip service to it, though modeling and explanatory practice in the discipline regularly, and increasingly, ignores it (Ross 2011). However, residues of methodological individualism, and the more general ontological atomism from which it derives, persist in influencing foundational assumptions. One of these is that coordination is sufficiently difficult in principle that it might furnish the basic challenge that was overcome by the evolution of humans' runaway evolution of cognitive plasticity.

No considerations reviewed in this chapter undermine, or are intended to undermine, the tight coevolutionary relationship between sociality and intelligence. However, the picture of initially isolated individuals under pressure to develop ingenious strategies for achieving coordination appears to be almost exactly backward. Primates and other social animals are equipped by basic and nonmysterious biological devices and behavioral dispositions to coordinate, at least in the statistical sense relevant to selection of mixed strategies without backward induction. But capacities for easy coordination are potential barriers to specialization of labor and to efficient exploitation of private information. The social evolution of norms that encourage and maintain individual variation among human selves is best understood in light of this.

NOTES

1. Lewis (1969) is usually cited in this connection. Lewis's account involved some significant distractions because he imposed the strong and unnatural constraint that a convention solves a coordination problem only in cases where all players are indifferent among the set of equilibria from which the convention picks a member. This restriction has its origins in the problems in the philosophy of language and the philosophy of science that

motivated Lewis's study, but it has no basis in either game theory (Ross 2008) or social theory (Gilbert 1989).

2. The standard normal-form version allows that the payoffs in each of the outcomes (Violence, Romance) and (Romance, Violence) might be (0,0). This is useful for ease of solution, and to illustrate a general logical point about randomization, but less intuitively natural than the payoffs given here.

3. From the very origins of game theory, modelers have attended to the strategic dynamics of players who can make binding promises to one another under the rubric of cooperative game theory (von Neumann and Morgenstern 1944). However, the main focus in cooperative game theory has been on coalition formation and has presupposed complete information.

4. Orangutans, which are among the most intelligent of mammals, were once thought to be solitary. This has turned out to be inaccurate as a generalization; and such solitude as is observed in some orangutans now appears to be a recent adaptation to habitat changes. See Dunbar (1988).

5. Subgame-perfect equilibria are those identified by backward induction in extensive-form games where all players have perfect information. In extensive-form games where some players are uncertain about the game structure, reasoning deductively from conjectures, and applying Bayes' rule to ensure consistent conditional probability assignments to such conjectures, identifies so-called sequential equilibria (Kreps and Wilson 1982).

6. Quantal response equilibria allow for varying indifference bands in agents' ordinal rankings of outcomes, and /or inability among players to discriminate among outcomes as finely as the analyst.

7. It can rightly be pointed out, following Deacon (1997), that human subcognitive coordination is not informationally restricted in this way because symbolic representation and linguistic processing structure representational dynamics throughout the brain. However, the point at issue here is that Thalos and Andreou's hypothesis can be defended without *needing* recourse to Deacon's.

8. Ofek argues that fire keeping was the first specialized occupation in human evolution. This rests on an economic argument that, for *H. erectus* and his immediate successors, it was much more efficient for specialists to maintain fires from which bands of local hunter-gatherers could draw in exchange for food and pelts than for each small band of hunter-gatherers to search for suitable kindling each day—which would have severely restricted their foraging ranges—and then endure the high-risk, failure-prone ordeal of starting a nightly fire without modern ignition technology. Thalos and Andreou object that Ofek fails to demonstrate that early humans could not have become trapped in the less efficient of these two production equilibria. This point is valid but irrelevant, since the internal logic of Ofek's argument merely requires him to show that the first equilibrium could have been arrived at by a sequence of behavioral adjustments that were all Pareto-improvements. This he does. Thalos and Andreou, in reconstructing Ofek's story, also ignore the importance in it of the scarcity of kindling. Leaving out this aspect makes the path to discovery of the first equilibrium seem much less likely to have been stumbled across than on Ofek's fully specified account.

9. Caves, Ofek argues, were not primarily used as homes by early humans, as popular imagination supposes, but as fire service stations. This naturally leads one to speculate, though he does not, that cave art might have had the intended function of attracting customers.

10. Bacharach (2006) is particularly clear and insistent in identifying this cause of intellectual frustration.

REFERENCES

Anderson, L., and C. Holt. 1997. "Information Cascades in the Laboratory." *American Economic Review* 87: 847–62.

Atkeson, A. 2001. "Rethinking Multiple Equilibria in Macroeconomic Modeling: Comment." In *NBER Macroeconomics Annual 2000*, B. Bernanke and K. Rogoff, eds., 162–71. Cambridge, MA: The MIT Press.

Bacharach, M. 2006. *Beyond Individual Choice*. Princeton, NJ: Princeton University Press.

Banerjee, A. 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics* 107: 797–817.

Barraclough, D., M. Conroy, and D. Lee. 2004. "Prefrontal Cortex and Decision-Making in a Mixed Strategy Game." *Nature Neuroscience* 7: 404–10.

Bruner, J. 1992. *Acts of Meaning*. Cambridge, MA: Harvard University Press.

Bruner, J. 2002. *Making Stories: Law, Literature, Life*. New York: Farrar, Strauss and Giroux.

Byrne, R., and A. Whiten, eds. 1988. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*. Oxford: Oxford University Press.

Camerer, C., E. Johnson, T. Rymon, and S. Sen. 1993. "Cognition and Framing in Sequential Bargaining for Gains and Losses." In *Frontiers of Game Theory*, K. Binmore, A. Kirman, and P. Tani, eds., 27–47. Cambridge, MA: The MIT Press.

Carlsson, H., and E. van Damme. 1993. "Global Games and Equilibrium Selection." *Econometrica* 61: 989–1018.

Corazzini, L., and B. Greiner. 2007. "Herding, Social Preferences and (Non-)conformity." *Economics Letters* 97: 74–80.

Deacon, T. 1997. *The Symbolic Species*. New York: Norton.

Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.

Donald, M. 1991. *Origins of the Modern Mind*. Cambridge, MA: Harvard University Press.

Dorris, M., and P. Glimcher. 2004. "Activity in Posterior Parietal Cortex Is Correlated with the Relative Selective Desirability of Action." *Neuron* 44: 365–78.

Dunbar, R. 1988. *Primate Social Systems*. London: Croom Helm.

Dunbar, R. 1998. "The Social Brain Hypothesis." *Evolutionary Anthropology* 6: 178–90.

Emery, N., N. Clayton, and C. Frith. 2007. *Social Intelligence: From Brain to Culture*. Oxford: Oxford University Press.

Gilbert, M. 1989. *On Social Facts*. Princeton, NJ: Princeton University Press.

Glimcher, P. 2003a. *Decisions, Uncertainty and the Brain*. Cambridge, MA: The MIT Press.

Glimcher 2003b. "The Neurobiology of Visual-Saccadic Decision Making." *Annual Review of Neuroscience* 26: 133–79.

Harsanyi, J. 1967. "Games with Incomplete Information Played by 'Bayesian' Players, Parts I–III." *Management Science* 14:159–82.

Herrnstein, R. 1961. "Relative and Absolute Strength of Response as a Function of Frequency of Reinforcement." *Journal of the Experimental Analysis of Behavior* 4: 267–72.

Herrnstein, R. 1982. "Melioration as Behavioral Dynamism." In *Quantitative Analyses of Behavior, Volume II: Matching and Maximizing Accounts*, M. Commons, R. Herrnstein, and H. Rachlin, eds., 433–58. Cambridge, MA: Ballinger.

Hollis, M. 1998. *Trust Within Reason*. Cambridge: Cambridge University Press.

Huck, S., and J. Oechssler. 2000. "Informational Cascades in the Laboratory: Do They Occur for the Right Reasons?" *Journal of Economic Psychology* 21: 661–71.

Humphrey, N. 1976. "The Social Function of Intellect." In *Growing Points in Ethology*, P. Bateson and R. Hinde, eds., 303–17. Cambridge: Cambridge University Press.

Hung, A., and C. Plott. 2001. "Information Cascades: Replication and an Extension to Majority Rule and Conformity-Rewarding Institutions." *American Economic Review* 91: 1508–20.

Kagel, J., and D. Levin. 1999. "Common Value Auctions with Insider Information." *Econometrica* 67: 1219–38.

Kreps, D., and R. Wilson. 1982. "Sequential Equilibria." *Econometrica* 50: 863–94.

Lee, D., M. Conroy, B. McGreevy, and D. Barraclough. 2004. "Reinforcement Learning and Decision-Making in Monkeys During a Competitive Game." *Cognition and Brain Research* 22: 45–48.

Lee, D., B. McGreevy, and D. Barraclough. 2005. "Learning and Decision-Making in Monkeys During a Rock-Paper-Scissors Game." *Cognition and Brain Research* 25: 416–30.

Lee, D., and X.-J. Wang. 2009. "Mechanisms for Stochastic Decision Making in the Primate Frontal Cortex: Single-Neuron Recording and Circuit Modeling." In *Neuroeconomics: Decision Making and the Brain*, P. Glimcher, C. Camerer, E. Fehr, and R. Poldrack, eds., 481–501. London: Elsevier.

Levitt, S., J. List, and S. Sadoff. 2011. "Checkmate: Exploring Backward Induction Among Chess Players." *American Economic Review* 101: 975–90.

Lewis, D. 1969. *Convention*. Cambridge, MA: Harvard University Press.

List, J., and D. Lucking-Reiley. 2002. "Bidding Behavior and Decision Costs in Field Experiments." *Economic Inquiry* 40: 611–19.

McKelvey, R., and T. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior* 10: 6–38.

Morris, S., and H.-S. Shin. 1998. "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks." *American Economic Review* 88: 587–97.

Morris, S., and H.-S. Shin. 2003. "Global Games: Theory and Applications." In *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress, Volume 1*, M. Dewatripont, L. Hansen, and S. Turnovsky, eds., 56–114. Cambridge: Cambridge University Press.

Noeth, M., C. Camerer, C. Plott, and C. Weber. 1999. "Information Aggregation in Experimental Asset Markets: Traps and Misaligned Beliefs." Working Paper: 1–50. Available at http://ideas.repec.org/p/clt/sswopa/1060.html.

Ofek, H. 2001. *Second Nature*. Cambridge: Cambridge University Press.

Richerson, P., and R. Boyd. 2005. *Not by Genes Alone*. Chicago: University of Chicago Press.

Ross, D. 2005. *Economic Theory and Cognitive Science: Microexplanation*. Cambridge, MA: The MIT Press.

Ross, D. 2007. "*H. sapiens* as Ecologically Special: What Does Language Contribute?" *Language Sciences* 29: 710–31.

Ross, D. 2008. "Classical Game Theory, Socialization and the Rationalization of Conventions." *Topoi* 27: 57–72.

Ross, D. 2010. "Naturalism: The Place of Society in Nature." In *The Sage Handbook of Philosophy of Social Science*, I. Jarvie and J. Zamorra-Bonilla, eds. Thousand Oaks, CA: Sage.

Ross, D. 2012. "The Economic Agent: Not Human, But Important." In *Handbook of the Philosophy of Science, v. 13: Economics*, U. Mäki, ed. London: Elsevier. Forthcoming.

Schelling, T. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Schweighofer, N., and K. Doya. 2003. "Meta-Learning in Reinforcement Learning." *Neural Networks* 16: 5–9.

Seabright, P. 2010. *The Company of Strangers*. 2d rev. ed. Princeton, NJ: Princeton University Press.

Searle, J. 1995. *The Construction of Social Reality*. New York: The Free Press.

Seo, H., D. Barraclough, and D. Lee. 2007. "Dynamic Signals Related to Choices and Outcomes in the Dorsolateral Prefrontal Cortex." *Cerebral Cortex* 17: 110–17.

Sgroi, D. 2003. "The Right Choice at the Right Time: A Herding Experiment in Endogenous Time." *Experimental Economics* 6: 159–80.

Skyrms, B. 1996. *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Skyrms, B. 2010. *Signals*. Oxford: Oxford University Press.

Spiwoks, M., K. Bizer, and O. Hein. 2008. "Informational Cascades: A Mirage?" *Journal of Economic Behavior and Organization* 67: 193–99.

Thalos, M., and C. Andreou. 2009. "Of Human Bonding: An Essay on the Natural History of Agency." *Public Reason* 1: 46–73

Turner, M. 1998. *The Literary Mind*. Oxford: Oxford University Press.

Von Neumann, J., and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Wang, X.-J. 2001. "Synaptic Reverberation Underlying Mnemonic Persistent Activity." *Trends in Neuroscience* 24: 455–63.

Wang, X.-J. 2002. "Probabilistic Decision Making by Slow Reverberation in Cortical Circuits." *Neuron* 36: 955–68.

West, S., C. El Mouden, and A. Gardner. 2010. "Sixteen Common Misconceptions About the Evolution of Cooperation in Humans." *Evolution and Human Behavior* 32: 231–62.

Whiten, A., and R. Byrne, eds. 1997. *Machiavellian Intelligence II*. Cambridge: Cambridge University Press.

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Blackwell.

Young, H. P. 1998. *Individual Strategy and Social Structure*. Princeton, NJ: Princeton University Press.