CHAPTER 11

From fitness to utility

Kim Sterelny

11.1 MODELLING AGENCY

This essay develops two themes. One is methodological, focusing on the role of model-based science in the human sciences, and, in particular, on model pluralism. I shall discuss three formally similar but causally distinct models of human agency, and shall argue that all three are important in understanding the interaction of individuals and their social context. Two are variants of the rational actor model derived from economics. The simplest of these is the classic rational economic agent model that treats agents as if they act to maximize their material wealth. I contrast it with another version, developed by Herb Gintis and others (Gintis 2006, 2007, 2009), that combines the formal machinery of game theory with the hypothesis that humans are strong reciprocators, and that the psychology of default co-operation plus revenge explains the uniquely co-operative nature of human social life. The third model is formally similar but different in both parentage and underlying causal assumptions, which derive from evolutionary biology. Human behavioural ecology sees humans as fitness rather than utility maximizers. This makes a difference, as we shall see when I take up the second and substantive theme of this essay: I aim to reveal the changing nature of individual agency in the transition from intimate to complex, stratified societies. I shall suggest that we need multiple models as a consequence of diachronic changes in the nature of agency, not just because of the complexity of individuals' relations with their social world. I begin by contrasting the Gintis model of rational agency with its ancestor, and then contrast both to a formally similar approach with roots in a different discipline, evolutionary biology. I then exploit these contrasts in exploring the changing demands on human decision making in the transition from simpler Pleistocene social worlds to the much more complex ones that followed in the Holocene.

Gintis (often in collaboration with Sam Bowles) has argued that humans can be modelled as utility-maximizing agents, thus keeping the formal and conceptual machinery of rational actor models of agency. But Gintis proposes reshaping the standard economists' version of that idea: the version that models humans as self-interested agents making optimal decisions about individual resource acquisition. Rational economic agents will co-operate when that is in their economic interest – you can do business with them – but they are never altruistic (or spiteful). Gintis follows Robert Frank (Frank 1988) in taking gene–culture co-evolution to have reshaped human psychology in ways that require a shift in modelling strategies. Humans have come to live in social worlds in which fitness depends on co-operative partnerships, and hence we have come to live in social worlds in which fitness depends on reputation. We reap the benefits of co-operative partnerships only if we are of good repute, for that makes it likely that we will be chosen by others for profitable partnerships. Fitness depends on agents seeming to be good, to be trustworthy. The best way of seeming to be good is to be good. Our moral emotions evolved in this selective regime, to be both signals of, and motivations for, our social other-oriented dispositions.

Gintis and his collaborators have enriched Frank's model by linking it to the experimental economics literature on 'strong reciprocity'; a set of experimental results that supposedly show that many humans enter interactions disposed to co-operate if they expect co-operation to be mutual, but also willing to punish free riders (Fehr and Gächter 2002; Fehr and Fischbacher 2003; Gächter and Herrmann 2009). Gintis and his colleagues have also developed a broader account of the psychological foundations of these co-operative dispositions. Gene–culture co-evolution has resulted not just in moral emotions but also in psychological predispositions to internalize norms and act on them; and many norms demand prosocial acts, even at individual cost. Finally, they show that this changed view of agency can still be accommodated within the formal framework of rational choice theory and game theory (Bowles and Gintis 2006; Gintis 2007). The crucial idea is that this changed view of agency can be accommodated by a more pluralist conception of agent utilities. Agents care about their own economic welfare. But we care, not only about our own welfare; we sometimes care about the welfare of others. If there are no constraints on what we can include in a model of an agent's values, rational actor models will be empirically empty. But these theorists show that a more pluralistic view of human goals is neither empirically nor computationally intractable, though it is messy.

What might this messier and more realistic conception of human motivation buy us? Realism need not be a virtue in itself. Rational economic agents are models rather than theories of human agency. Models (in contrast to theories) represent their real-world target systems indirectly.[1] A model is not a partial, incomplete or idealized direct description of the world: it is a fully accurate description of an ideal or fictional system. Model specifications are not schematic or partial descriptions of real-world systems: such specifications do not just suppress irrelevant detail; they specify properties known not to be satisfied by any real-world system. So, for example, many population genetics models are models of infinite populations within which organisms mate at random. No biological populations have such properties. Models are useful to the extent that these ideal systems are importantly similar to real-world target systems, and we can use those similarities to give us explanatory and predictive leverage over real-world systems (Godfrey-Smith 2006; Weisberg 2007; Godfrey-Smith 2009). So, for example, the famous prisoner's dilemma is not a realistic picture of real human interactions. We do not make our choices simultaneously, in complete ignorance of others' choices, with perfect knowledge of the outcomes of those choices, and with those outcomes being fully quarantined from all other social interaction. Even so, it is widely accepted that models of iterated prisoner's dilemmas capture important causal mechanisms in human co-operation and its breakdown. Models can also be very informative when we identify their limits: when we can identify when and why the model–target system similarity breaks down. Again, the prisoner's dilemma literature provides an illustration: these models very clearly show that stable multilateral co-operation requires a different explanation than stable bilateral co-operation.

So there is no point in complaining that unlike rational economic agents, real humans care about more than economic resources, or that we are imperfectly rational in assessing odds, given our information. Model specifications are never true of any real-world system. The crucial question is: in what respects do human agents resemble rational economic agents, and what are the consequences of those resemblances for human social worlds? There is much controversy in philosophy of science about when, and why, a model-based approach to science is appropriate (Wimsatt 2007). But there is a partial consensus that (i) models are important when

[1] In developing this view on the role of models in an integrated human science, I have been heavily influenced by John Matthewson, Brett Calcott, Michael Weisberg and, especially, Peter Godfrey-Smith, whose picture of the role of models in science I have taken over.

we are confronted with domains that are complex and causally entangled, for in such cases direct representations are computationally intractable. Models represent target systems through similarity relations between the model and the target. Since models can be similar to targets in different respects, we can consistently use distinct, inconsistent models of a single system. For each model can be similar to the target in different ways. (ii) Models are important exploratory tools in the earlier stages of a discipline's development. For theorists need not specify in advance model–target system similarities. One research strategy is to explore the properties of a model (or of a class of related models) and look for surprising similarities between model dynamics and those of the target system.

I shall explore a version of this strategy in this chapter. Human social worlds *are* complex and causally entangled. But they are also dynamic. For consider how deep human history is, and how profoundly the demographic, social and economic worlds of humans have changed over the last 10,000 years. One hundred thousand years ago, *Homo sapiens* individuals lived in small foraging groups, probably confined to Africa, with technology still limited in its diversity and still changing very slowly. Hierarchy and social complexity were probably minimal, with specialization based on division of labour still in the future. Yet these ancient humans were certainly cultural and social beings. Their technology was limited in its variety, but its production and effective use demanded skill, knowledge and co-operation. Over the past 10,000 years, the geographic, demographic, economic and social worlds of most humans have changed in ways those ancient humans could not imagine. Informational demands have changed, as human social worlds have become larger and more stratified, with human action often being routed through social institutions with specific routines and protocols. The time horizon of decision and planning has lengthened: foragers act in time horizons of hours, days, perhaps a week. Subsistence farmers invest time, labour and energy in activity that will pay off, if at all, months later. Contemporary humans invest in education and pension plans whose payoffs come decades after investment. There are important changes in the organization of motivation, as well. Norms, formal rules, and actions based on institutional role replace or supplement motivation based on direct bonds of affiliation and aversion. Paul Seabright, in particular, has emphasized that the Holocene world increasingly involved trusting interaction with strangers, a major break from the past. The extent of these changes (I shall suggest) makes it likely that we need more than a single model of agency. The nature of agency has changed, as human social worlds were transformed.

In particular, I shall suggest that a dynamic view of agency makes sense of two co-existing approaches to agency. One set of ideas derives from economics and allied disciplines, disciplines which represent agents as effective instrumental reasoners. Human behavioural ecologists, in contrast, have adapted tools from behavioural ecology to model humans as fitness optimizers. Human behavioural ecology seeks to show that apparently puzzling patterns of human social worlds – for example, apparently wasteful food taboos – are in fact consequences of adaptive decision making, once we understand the options agents have, and the costs and benefits they must balance. The formal machineries of utility-maximizing and fitness-optimizing models are very similar. But the quantities optimized are not, and nor, as we shall see, are the mechanisms that explain how actions are shaped to ends.

11.2 ECONOMIC AND OTHER AGENTS

In this section and the next I explain the two modelling approaches and their strengths. I then suggest a dynamic that explains the transition from agents best modelled through the lens of behavioural ecology to those best modelled with the tools of economics. It is the economist's picture of agency with which I begin. As noted, rational economic agents care only about their own individual welfare. While this simplifies the variety and idiosyncrasy of human motivation, this is an advantage of these models rather than being a weakness. These simplifying assumptions about motivation give them real predictive and explanatory power. For the simplifying assumption is not arbitrary: real humans might care about more than their own material welfare, but they do care about their welfare, typically according it a high priority. Moreover, the assumption that agents have self-regarding economic preferences enables models to be general, strategic, empirically tractable, and explanatorily salient.

The framework is general, because problems of resource acquisition, and the problem of managing trade-offs between risks and benefits, arise in all human social worlds. Rational economic agent models have their most direct application to markets. But they have been applied to a raft of other resource allocation contexts as well, for example, to explanations of the effect of urbanization on family size. Rational economic agent models are consistent with the interactive, strategic character of individual decision making. In many contexts, the consequences of my choices depend in part on others' choices: if I expect you to block an escaping stag by going left, I should stay put; but if I think you will stay put, I should break left. While we do not always choose well (by our own lights) in strategic contexts, we often do, and that shows that we are not opaque to one another: we must often be able to estimate others' goals and their best path to those goals. Rational economic agents who know that they are interacting with other rational economic agents will typically be translucent to one another. The agents in a local interaction will often be able to estimate the local resource envelope that is available, the relative worth of different goods (which will in part be a function of their scarcity and/or the expense of their production), and the most effective resource acquisition strategies. Their own economic activities will generate as a side effect useful information about what other agents want, and how they might get it. So rational economic agents will often make a decent fist of strategic interactions, just as we do.

Rational economic agent models are empirically tractable. There will always be some doubt about whether a given action has maximized the possible gains. For (first) actions are optimal only given a specified set of alternatives, and (second) the fact that an action happened to turn out well (or badly) might show luck rather than best choice. But unlike subjective utility, the economic returns to action are public. So actual outcomes are measurable in the same units the agent uses to value outcomes. Moreover, since models are constructed to represent decision and outcome in reoccurring situations rather than idiosyncratic ones, in practice it is possible to discriminate between good luck and good choice.

Finally, the appeal to economic utility is explanatorily salient. The idea here is that we rob rational actor models of explanatory power if we incorporate norms and customs, and especially culturally contingent norms and customs, into the agent's conception of value. The importance of enculturation is undeniable, but culturally specific norms should not be incorporated into agent preference structure to explain action. We cannot explain (say) the gerontocentric distribution of sexual access in many traditional aboriginal communities (see Keen 2006) by supposing younger male and female agents to have a normatively shaped preference for such partners. For the origin, spread, persistence and influence of such norms is a crucial *explanatory target*, not an *explanatory resource*. Similarly, the East Asian preference for boy children, even in an environment of female shortage, as revealed in patterns of infanticide, is an explanatory target. In these puzzling cases, we would learn nothing from a model of individual choice which simply factored into agent utility functions a strong preference for boys over girls, even when there is independent evidence that agents do have such preferences. Thus a causal model that depends essentially on the motivational power of norms is explanatory when and only when it includes an independent explanation of the origin and stability of those norms.

Thus the ruthless simplicity of the rational economic agent model is crucial to its power. The key question is whether this framework successfully accounts for human co-operation, given realistic specifications of the size of co-operating groups in human cultures, the levels of co-operation they achieve, the frequencies with which individuals expect to interact, and the costs of directing punishment at free riders. As group size increases, a rational economic agent is less likely to co-operate to secure the benefits of repeated direct reciprocation. For in larger groups, when all else is equal, any two agents are less likely to interact regularly. But perhaps in the right circumstances, even in larger groups, self-interested agents will co-operate because reputation is an important resource. It can pay Alex to help Alfred, even if Alfred is unable to return the favour, and even at some cost. For others may see Alex's action and use it as a cue for positive discrimination in future interactions. A self-interested Antonio might use Alex's helpful act as such a cue, both because Alex has signalled readiness to co-operate and (once this show gets off the ground) because Antonio's own reputation will be improved by his positive treatment of other co-operators.

Gintis, Bowles and many others are profoundly sceptical of this vision of the emergence of co-operation among self-interested, reputation-sensitive agents (there is a particularly illuminating exchange between Gintis and Paul Seabright on one side and Don Ross and Ken Binmore on the other: Binmore 2006; Gintis 2006; Ross 2006a, 2006b; Seabright 2006). Their critique has two elements. First, they argue that there are experimentally constructed target systems – experimental public goods games, ultimatum games, three-player ultimatum games, and the like, which are demonstrably not captured by reputation-sensitive rational economic agent models. For (i) these target systems are structured to exclude the possibility of rationally investing in reputation; and (ii) in an important class of cases, co-operation still emerges. Moreover, these targets are structured in such a way that if agents have non-self-interested, social utilities, there will be a clear and consistent signal of those social utilities. Such signals are indeed seen.

Second, they argue that these models of reputation-mediated co-operation do not capture crucial aspects of co-operation as it emerges in natural communities: there are crucial differences between the fictional systems of interacting, reputation-sensitive but self-interested agents, on the one hand, and these real communities, on the other. In the models, agents are unrealistically well-informed about one another. In the models, reputation tracks actual action too exactly, because in real life, social information is not free. Thus an agent refusing to co-operate with a third party might be defecting, or might be punishing a defector: the difference is important; but it is not obvious to casual inspection, and not available for free. So real agents' information about their social constellation will be both incomplete and somewhat inaccurate. Moreover, these models of prudential co-operation assume that punishment – the exclusion of free riders through negative discrimination – is virtually free. But that is not true of target systems with high levels of co-operation.

11.2.1 What does experimental economics show?

In response, defenders of the rational economic agent framework have argued that the experimentally constructed target systems of behavioural economics are themselves models of naturally occurring, socially salient real-world systems,[2] and they are poor models of such systems. In developing this sceptical view of the experimental data supposedly showing that humans are 'strong reciprocators', it is important that defenders of the rational economic agent framework do not take that framework to represent the computational architecture of the mind. Rather, they are models of resilient behavioural dispositions in a central class of decision contexts. These contexts are (a) central to the life prospects of the agent in question and central, through the outcomes of decision, to the social life of the community of which the agent is a member; and (b) stereotypical.

The agent in question (and typically other agents in similar situations) has a history of decision making in similar situations, with similar options and cost–benefit trade-offs. In virtue of the first of these conditions, it is worth developing effective decision heuristics; in virtue of the second, it is possible to do so. From the perspective of cognitive science, we are probably far-from-ideally rational, mixed motivation bundles of kludges. But with appropriate practice, and with appropriate environmental support, in a limited but salient set of contexts, we can marshal heuristics that approximate ideal rationality. In many experimental economics settings, these heuristics do not deliver economically optimal

[2] But a model not in the sense of a formal or fictional model, but a model system: in the way rats and pigeons have been model systems for those working on associative learning, or the fruit fly has been for those working on genetics and development.

outcomes. For these experimental contexts have been designed to subvert adaptive decision heuristics, to drive a wedge between what is prudent in reputation-sensitive, repeated-interaction contexts and what is prudent in blind, often one-off contexts. We rely on habits and values which are prudent in many naturally occurring and familiar settings, but these will often not optimize in many experimental settings. Even though they fail to optimize in such experimental settings, perhaps most humans act like rational economic agents in the core business of their daily life, and that of their community. If so, that is vindication enough of rational economic agent modelling.

From this perspective, whether an agent is well-modelled as a rational economic agent does not depend on the subjective, introspectable character of their immediate motivation. It depends on whether in the socially central contexts these motivations – independently of their qualitative or representational character – in fact reliably prompt action that advances the material interests of the agent. In experimental contexts, the 'pro-social' emotions identified by Fehr and Gächter may well lead to a sacrifice of material interest. It by no means follows that they are similarly sacrificing in real-world target contexts. The experimental games provide evidence that the rational economic agent framework does not model the intrinsic computational architecture of humans well. Hence it does not give a context-general model of human decision making. But few defenders of the rational economic agent framework think of it as a model of *thinking*. They see that framework as modelling the outcomes of choices in strategic, repeated, highly salient contexts. And that idea remains live in the face of experimental economics.

11.2.2 Would the prudent but selfish co-operate in multiplayer interactions?

The core social business of human life includes many co-operative, multi-agent interactions. If these fall outside the scope of rational economic agent models, these models are brutally limited. Ross and Binmore argue that Gintis, Seabright and others overstate the problem of information access and expensive punishment, problems that supposedly undermine prudential co-operation in such interactions. They point out that information flows through a community as a side effect of agents' utilitarian activities. Individuals interact, and observe others interacting, in the course of their ordinary activities. As individuals interact, they leak information about their capacities, habits and dispositions. To find out about Alex and

Alfred, often Antonio does not have to invest in epistemic action; he simply needs to be ready to pick up freely available information. (Such free information is not full information, but it may be full enough.) Moreover, they argue that punishment is typically costless too. In most circumstances, punishment consists of no more than a zero-cost social warning ('I see you') at an incipient social transgression, a signal that is noted and acknowledged as the transgressive act is aborted with no damage done. In support of this interpretation, Don Ross sketches a familiar pub scenario: an agent who seems reluctant to pay for his round is reminded of the standard operating procedure with a just-pointed-enough joke. The incipient freeloader publically takes it as a joke that includes rather than excludes him. But he buys his round, and social amity is restored. Policing is necessary, effective, but because it is pre-emptive, heading defection off rather than punishing after it has occurred, it is cheap.

In my view, Ross and Binmore are largely right about information, but wrong about punishment; once co-operation is established in human groups, information does flow freely, because work is social. We extract, transform and often distribute resources in teams. Close ecological and economic co-operation generates information about one another as a free side effect (Sterelny 2007). Once co-operation is established as a stable feature of human social worlds, agents will often no longer routinely have to invest time and effort in finding out about one another (though in special circumstances, in high-stakes decisions, they may). However, this free access to information changes in the transition to larger and more complex societies in the late Holocene. As social worlds become larger, and as they develop more specialization and social stratification, the automatic flow of information in the village glasshouse dwindles, and increasingly we need to rely on third-party opinion through gossip networks, and on social markers (the right accent, the right clothes). The cost of information goes up, and its reliability falls.

Even so, we do know a lot about other agents. I am more sceptical of their line on punishment. Ross's scenario is indeed familiar, and he is right in thinking that pre-emptive punishment can be low-cost and effective, for no costs on co-operators have yet been imposed by defection. But it is a best-case scenario. It ignores risk costs. Those of us who share Ross's familiarity with norm management at the pub will also recall social groups fracturing, as one agent resents attempted social sanction[3]

[3] Often because they do not think they have violated any norm that they recognize – there does seem good evidence that we much resent mistaken punishment.

and tempers flare. At the limit, such irruptions can be lethal (Seabright (2010) details the very high murder rates in co-operative forager societies). Attempted low-cost punishment always risks turning into high-cost, even very high-cost, punishment. Moreover, we need to ask why 'soft' social sanctions are often effective. In part, they are effective because of an implied threat to exclude the agent from a material benefit. Even a rational economic agent may weigh heavily a threat to the flow of beer. But they are also effective because people care about subjective interpersonal rewards, about how others see them. Rational economic agents are relatively impervious to soft sanctions, for they have only an instrumental interest in others' regard.

The strong-reciprocation conception of human agency can be partially reconciled with the rational economic actor model. It may well be that in the intimate world of the Pleistocene foraging band, the cool-hearted, farsighted prudent rational economic maximizer would have co-operated consistently with others in the community. Co-operation was prudent: there were very considerable staghunt and risk-management gains from co-operation. Free-riding would be imprudent: interaction was frequent; information was widely available. Punishment was sometimes cheap (for example, through exclusion in contexts of partner choice). When not cheap, punishment could sometimes be co-opted as an honest signal of quality and commitment. So free riding would be risky. Rational economic maximizers in the Pleistocene might well look just like strong reciprocators. Perhaps that is how our ancestors managed maximization. For as Sarah Hrdy (2009) very vividly illustrates, we were not, and did not evolve from, cool-headed calculators but from hot-headed great apes. So the psychology of strong reciprocation may well have been the only available way of engineering an approximation to rational economic agency from a starting point of hominins with rudimentary but powerful social emotions and poor impulse control. But once the flat and intimate social world of the Pleistocene was displaced by stratified societies in the Holocene, the two models of agency diverge. They probably predict very different patterns in action, as opportunities for free-riding emerge.

Strong-reciprocator psychology is still a version of instrumental rationality. Strong reciprocators have utility functions in which objective costs and benefits are modulated by subjective costs and rewards, as the welfare of others counts to an agent in his or her assessment of the utility of an act. For example, punishment is contingent on perceived cost and impact. Thus if competition is added to the ultimatum game, so the first player can make an offer to several candidate second players, those second players are less willing to punish a low-ball offer by refusing (presumably because the punishment might fail through the low-ball offer being accepted by the rival second player). Strong reciprocators do not punish independently of costs and benefits. So this version of the rational actor model has many of the virtues of the rational economic agent framework. Agent utilities are common across agents and stable over time, for these depend on social and moral emotions that are developmentally stable features of our evolutionary inheritance. As with the rational economic agent model, this view of agency fits with our capacity to manage strategic interaction. We are translucent to one another, because our utilities in part depend on our access to objective, public resources, and in part because they depend on subjective features of agent psychology which are displayed by emotional signals. Importantly, to the extent that these rational social agent models are based on prosocial emotions rather than norm acquisition, social preferences are independent of, and explanatorily prior to, the social phenomena – the networks of human co-operation and pro-cooperation norms – that these models explain. Prosocial emotions are an explanatory resource for a theory of human co-operation (granted a theory of their evolution). They are not just one of the phenomena such theories are meant to explain. Finally, rational social agent models have broad scope, as they are meant to help explain not just a wide range of co-operative interchanges, but also those in which deterrence is important, for example bargaining interactions where the weaker side must have a credible option to walk away from the negotiations. In these models, the social and moral emotions make our threats credible, not just our promises.

The idea, then, is that this model of agency is not as simple as the rational economic agent model, but it will predict human action in a broader range of contexts. As noted above, that may be crucial. For while rational economic agency models might well predict that humans co-operate in the intimate glasshouse of Pleistocene forager bands, it is much less clear that they predict co-operation in the larger, noisier and less intimate worlds of the Holocene.[4] The bottom line: rational actor models can be empirically tractable, general and explanatorily powerful without presupposing that agents have self-interested motivations.

[4] And yet the social contract did survive in these larger social worlds, even before the establishment of coercive machineries of early states. I take up this puzzle in Sterelny (forthcoming).

11.3 HUMAN BEHAVIOURAL ECOLOGY

There is therefore an impressive case to be made for rational agent models of human action. However, co-existing with these economics-derived models, there is an alternative framework that descends directly from evolutionary biology. Animals sometimes behave in very puzzling ways, engaging in expensive displays, long migrations, sometimes imposing limits on their own seasonal reproduction, sometimes reproducing once, and then dying. Some change sex, others mimic the opposite sex. Behavioural ecologists have developed plausible models showing that such apparently dysfunctional behaviours in fact maximize fitness. Thus behavioural ecologists represent organisms as fitness optimizers, though they optimize subject to constraints imposed by their potential phenotypes, their environment, and the choices of other agents (Krebs and Davies 1997).

Human behavioural ecologists have co-opted this program. Human behavioural ecologists build models of distinctive, central patterns of human action: fertility decisions, food gathering, resource sharing. Individuals optimize. But they optimize subject to constraints imposed by the conditions under which they act and the choices of other agents (Winterhalder and Smith 2000; Laland and Brown 2002; Smith and Winterhalder 2003; Laland 2007). Such models of adaptive decision making have a common structure. They specify the optimal behaviour, given a set of assumptions about the environment, the range of options available, and the impact of choice on fitness. Forager women, for example, face difficult trade-offs over birth spacing, as their nomadic life style and the costs of mobility mean that they cannot care for a second child until the first can travel unassisted (see for example Kaplan 1996; Blurton Jones et al. 2006). So a !Kung pattern of birth spacing is the result of choice constrained by social and physical environment. !Kung women respond both to the objective resource envelope and to the choices that others make. If her social group moves, a !Kung mother must move with it (Blurton Jones 1997a, 1997b).

There is considerable predictive overlap between these three approaches to human behaviour. Seeing humans as strong reciprocators and seeing them as rational economic agents make equivalent predictions when individuals maximize their resources through co-operation, and where it is prudent to control defection by punishment. In turn, these models are equivalent to those from human behavioural ecology in those contexts in which material wealth is the most crucial fuel for fitness. But these are substantive restrictions, so the overlap is partial. There is an obvious question, for example, about the extent to which maximizing wealth maximizes fitness in contemporary western societies. There is a similar issue about forager societies. Eric Alden Smith and his colleagues argue that embodied and social capital are more important to forager success than material wealth, in part because they have little opportunity to accumulate material wealth and in part because skill and social support are so important to their lifeways (Smith et al. 2010).

An impressive and illustrative paradigm of human behavioural ecology is the long debate over male large-game hunting in forager societies (recently reviewed in Gurven and Hill 2009). On one side, it has been argued that male hunting cannot be explained as an essentially economic activity: because it is not profitable enough relative to alternatives or because hunters do not keep enough of these profits. According to this view, hunting is a form of signalling. Their critics argue that the signal hypothesis underestimates the profits of hunting and the extent to which sharing those profits is contingent on return benefits (Bliege Bird et al. 2001; Hawkes and Bird 2002; Gurven and Hill 2009). Once we do the accounting correctly, hunting makes economic sense. Despite these differences in interpretation, both sides presume that agents typically maximize their individual fitness (or perhaps that of their family or local group). Large-game hunting is a long-established, widespread, core activity in many human cultures. Such patterns of action reflect and respond to human selective environments. Neither those that defend the signalling model, nor those sceptical of it, think male hunting is satisfactorily explained by the existence of norms that endorse both hunting, and sharing the proceeds of those hunts, even though such norms exist (Boehm 1999). Their shared assumption is that norms inconsistent with interest are unlikely to establish or be stable, if established.

Human behavioural ecologists portray agents as acting adaptively, given their environments and the options open to them. The idea, though, is not just that some forms of human action are adaptive. Rather, these forms of action occur *because* they are adaptive. There is a systematic connection between environmental challenge and agent response: if we understand the challenge, we can predict and explain the response. How, though, does the fact that (say) hunting success enhances fitness by signalling good genes explain the fact that men hunt? They are not consciously aiming to maximize the number of their grandchildren through such signals. Nor is male hunting a wired-in, innate drive. Few males in contemporary cultures impress women by stopping for roadkill. The use of the same formal models of standard behavioural ecology disguises the fact that the explanatory structure of human behavioural ecology is quite different from that of its parent discipline.

In behavioural ecology simpliciter, adaptive fit depends on population history. Salmon spawn and die because they are playing a game against the world, and given its parameters, they do best by one-shot breeding. River journeys are high-risk activities for large, visible fish in small clear streams. There is no point in saving resources for the future if there is unlikely to be a future, and so salmon maximize their reproductive success by staking all their resources on the first mating effort (Quinn 2005). Ancient salmon that played a different strategy – that reserved resources for a second attempt – were less fit than those who invested more in their first go. The result is a population with little variation or individual flexibility. Salmon would not (I suspect) respond to safe streams by staggering their reproductive effort. The adaptive fit between action and environment depends on the history of the population, as it experiences repeated selective filtering. In contrast to the models of agency that descend from economics, adaptation depends on the 'intelligence' of the lineage, not just that of the individual agent.

The salmon breeding strategy is invariant. But even when populations vary, adaptive responses studied by behavioural ecology often depend on population history. Consider, for example, those species in which we find multiple mating strategies. Powerful males guard resources; other males seek sexual access by mimicking females (Jukema and Piersma 2006). These systems often include individual flexibility: as a male matures, he may switch from a cryptic to a resource-defending strategy. But even though the animals act in different ways in different stages of their life, they continue to conform to a single, though conditional, behavioural rule. Each male cuttlefish (for example) conforms to the strategy: guard if you are large; otherwise imitate being female (Norman et al. 1999). As with salmon investment decisions, this strategy is pervasive in the population as a result of selection on previous populations which probably included other mating strategies. Perhaps some males attempted to defend resources independently of their size; others did not switch when the time was right. The adaptive response of individuals is explained by the repeated filtering of variation that existed in previous populations.

Rational agency models depend on individual adaptive response in explaining the fit between an agent's actions and goals. Human behavioural ecology is intermediate between classic behavioural ecology, where fit depends on the evolutionary plasticity of the lineage, and the intelligent, informed, adaptable individual of economics. Population-level processes do play a central role in explaining human fitness-maximizing strategies. But the models of human behavioural ecology typically presuppose individual adaptive phenotypic plasticity. Let me illustrate its role before turning to that of population-level processes.

Consider a classic of the 'hunting is a signal' literature: Bliege Bird et al. (2001) on Meriam turtle hunting. They argue that turtle hunting is a costly signal of male fitness: its rewards are those of status and esteem, rather than calories. The fitness model may be the same as that of Zahavi on the Arabian babbler (Zahavi 1990). But there is no assumption that the strategy of hunting turtles as a signal has outcompeted other strategies over deep time. The Mer live partly by subsistence and partly in the cash economy, and hunting turtles for status might well be a quite new custom; certainly, they use store-bought equipment in the hunt. Turtle hunters may well be maximizing their fitness, but individual adaptability plays a central role in establishing a match between actor, environment and reward. Human response to the environment owes everything to behavioural plasticity. We adapt through learning, and those capacities often enable agents to adapt in spite of new challenges from the physical, biological and social environment.

As with rational agent models, human behavioural ecologists do not model thinking, certainly not conscious, explicit thinking. We do not consciously, deliberately, maximize fitness. So rather than being models of individual computational architecture, they too are models of resilient behavioural dispositions. The core achievement of these models, when they are successful, is to identify the real options available to the agent, and the costs and benefits imposed by the environment on these options. So these models represent the interaction between choice, current environment and outcome. But while they do not model thinking, they have cognitive presuppositions. They implicitly rely on agents having the ability to track their world. The explanation of !Kung fertility relies on humans having cognitive mechanisms that allow them to assess the causal structure of their environment, and to recognize the likely consequences of their actions. And it presupposes that proximate motivation tracks fitness. Humans do not consciously aim at maximizing their fitness any more than salmon do. But the outcomes we prefer increase fitness; those we avoid reduce fitness. So while !Kung women do not plan to maximize the number of their surviving grandchildren, they do plan to have (say) well-fed and well-behaved children, and those goals (if achieved) covary systematically with their fitness. The proximate targets of action are fitness resources. As we shall see, population-level processes play a key role in stabilizing the connection between these proximate goals and fitness resources, and in providing the informational resources for adaptive action.

11.4 FROM FITNESS TO UTILITY

Human behavioural ecology models have been very insightful in exploring dynamics of traditional cultures. So, for example, there are models of human life-history evolution and the puzzling fact that despite human infants being extraordinarily helpless, expensive and long dependent, we have relatively short interbirth intervals compared with chimps. There are 'polygyny threshold' models that explore the circumstances in which a women should choose to be a second wife in a multifemale household (Marlowe 2000). Yet while these models have been widely and insightfully applied to human life in small-scale traditional societies, these models have been almost invisible in explaining human action in large-scale worlds (perhaps with the exception of mate choice). Why would that be? If we can insightfully model agents as fitness maximizers in small worlds, why is it so much less useful for large worlds? That is not because small-scale worlds are precultural, that 'biology dominates culture in, but only in, such worlds'. Ethnographers have amply documented the very rich cultural life of historically known foragers and traditional agriculturalists. While there is some controversy about the richness of the symbolic and normative life of the very earliest members of our species, the physical record of rich culture is at least 40,000 plus years old (Klein and Edgar 2002). Moreover, natural selection continues to act in large-scale social worlds. Indeed, if Jared Diamond and others are right, the fate of large-scale social worlds has often depended on differential susceptibility to different diseases (this idea is not new; see Zizsser 1935, Diamond 1998, Crosby 2004). All core human activities are profoundly influenced by both our inherited culture and our inherited biology.

So at this point of this essay, the emphasis changes from the first to the second theme: from methodological issues about models and model choice to the substantive project of revealing the changing nature of individual agency in the transition from intimate to complex, stratified societies. The transition to large-scale social life reshapes the sources of cultural information. Individuals remain adaptable and responsive, but population-level mechanisms change, reshaping the extent to which individual decision making tracks fitness. As a consequence of reshaped cultural learning, proximate motivation is less well tuned to fitness interests. In fluent, practiced decision making, we continue to make effective decisions to get want we want. But what we want less reliably tracks our genetic interests. The propensity of individuals in small-scale worlds to make near-optimal fitness-maximizing decisions depends on their capacity to acquire the information they need and to tune their subjective goals to their objective needs. These mechanisms of individual adaptation are stressed in the transition to large-scale social worlds. As human social worlds shrink in number and variety, but grow in size and vertical complexity, the mix of biological and cultural factors changes. As a consequence of that changed mix, human agents shift from being fitness maximizers (perhaps roughly approximated as resource maximizers) to utility maximizers. For the transition to mass society makes tracking instrumental information more complex, and makes the connection between proximal motivation and fitness more fragile. But despite the increased complexity of these new social worlds, our mechanisms of rational appraisal are intact. In a range of key cases, we continue to make instrumentally effective decisions. But the link between conscious goals and fitness resources is fractured. We less reliably want what our genes need.

11.4.1 Population-level mechanisms

Population-level mechanisms are important in explaining the link between decision making and fitness in traditional cultures. Vertical (and near-vertical) cultural inheritance is important for the accumulation of cognitive resources essential for instrumental rationality. Boyd and Richerson are fond of contrasting the grim fate of the Burke and Wills expedition across central Australia (almost all died) with the untroubled survival of the locals, who had the benefits of the accumulated lore of their ancestors. We live in complex and somewhat conflicted social worlds; we often live in risky physical environments; we live by extracting high-value but often heavily defended resources from the biological and physical environment. So effective instrumental reasoning is information hungry. In small-scale human social worlds, this information is assembled gradually, generation by generation, filtered, gradually improved, and transmitted to the next generation with high reliability. Peter Richerson, Robert Boyd and their co-workers have shown this process to be central in guiding adaptive behaviour in traditional social worlds. They developed an array of formal models of transmission, mostly to show that reliable transmission may not require high fidelity learning in specific learning episodes. Redundancy and repetition can compensate for imperfect learning. Yet to make adaptive decisions, agents need more than good information. They need good goals too. Since agents do not (typically) subjectively aim at maximizing fitness, human behavioural ecology presupposes that there is a stable correlation between the proximate desires of agents and fitness: what agents want are fuels for fitness. Children inherit values as well as information from their parents. Maladaptive values are filtered out, just as maladaptive beliefs are.

The transitions to mass society make these filtering mechanisms less effective: children acquire less of their information and less of their values and goals from their parents, and from close associates of their parents who are informational and ideological duplicates of their parents. For as social worlds become larger and more connected, the shape of the cultural transmission network has changed from closed-vertical flow to open, much more oblique and horizontal flow. As social worlds become larger, individuals resemble one another less and less in belief and value: for these reflect not just idiosyncratic individual history but social and occupational role, and place in the social hierarchy. So the difference between vertical and other forms of cultural learning makes an increasing difference to what is acquired. So a hypothesis suggests itself: with the transition to mass societies, there is a radical decline in individual-level heritability of those traits whose development is culture dependent. In traditional societies, children ideologically and informationally resemble their parents, because they inherit – to a significant degree – their values and goals, and their instrumental lore. Seriously maladaptive lore and values are less likely to be rebuilt in the next generation, for the same reason seriously maladaptive genes become rarer.

So individual to individual heritability may well have declined, with increasing social complexity. That is not the only important change in population-level processes. There is a persuasive case for the idea that selection on culturally defined human groups has been a mechanism of potential importance in human evolution. Group selection is powerful when (i) the metapopulation is large, with significant and stable variation between groups; (ii) groups differ from one another in their prospects for survival, and for founding new groups; (iii) groups that bud from parent groups resemble those parents; (iv) the groups themselves are internally homogenous (for otherwise individual selection within groups may preempt group selection). Arguably, the growing importance of culture and cultural life resulted in human metapopulations satisfying these conditions (Sober and Wilson 1998; Richerson and Boyd 2001; Richerson et al. 2003; Boyd et al. 2005; Gintis 2007). Group selection would reinforce individual cultural inheritance: groups which have practices that make them more efficient in accumulating, filtering or transmitting high-utility information would be favoured at the expense of groups less able to accumulate or conserve their informational resources.[5]

Arguably, cultural group selection has played an even more important role in filtering maladaptive norms and values: in ensuring that what people value is good for their reproductive prospects. Gintis makes much of the fact that (most) people internalize the norms of their community: they make the community's values their own, and that influences what they want. Thus, many in the west find the idea of eating dogs not just contrary to the prevailing social norms but repellent, disgusting. Boyd and Richerson point out that there is no local guarantee that the norms established in a community are adaptive for that community or the individuals in it. There need only be a small fraction of zealots to establish a local custom, for they are willing to punish those who do not conform (Boyd and Richerson 1992; Boyd et al. 2005). Once established, it then becomes normalized for the succeeding generations. If the zealots are zealous enough, seriously maladaptive norms and customs can become locally entrenched. Boyd and Richerson suggest that such maladaptive norms are weeded out by group selection. Communities with maladaptive norms fail to prosper in competition with communities whose norms and customs encourage activities that are adaptive for the group (and/or the individuals within it).

The importance of group selection to human evolution remains controversial. But it may well be that adaptive customs in traditional societies depend in part on such selection. We see customs that help preserve expertise because groups that were poor at protecting their cognitive resources have disappeared, along with their members. Likewise groups whose norms encouraged maladaptive ends are no longer around to trouble ethnographers. We see adaptive action because of group-level filtering of communities with maladaptive practices. Even if this mechanism was once important, it can no longer be powerful. The transition to mass society has decreased metapopulation size, and has increased the internal heterogeneity of the remaining populations. These changes depower cultural group selection. So cultural group selection will filter norms and beliefs much less effectively. To sum up the argument so far: heritability declines, and cultural group selection is less powerful. So adaptive

[5] Perhaps, for example, there was selectable variation in local practices that support the division of informational labour and specialization – hence accumulating information more efficiently – or practices that support more effective teaching, for example by according high prestige to those prepared to share their expertise.

responses cannot be sustained by population-level mechanisms; lineages are less 'intelligent'. In mass society, adaptive action will depend on mechanisms of individual adaptive plasticity, plasticity of belief and of value.

11.4.2 Instrumental rationality in complex societies

Enculturation increases the informational load on adaptive action. Compared with our minimally cultured ancestors, wielding something like Oldowan technology, and living in social worlds roughly comparable to those of chimps in size and complexity, later hominins needed a much richer stock of information. A brief and partial list of the increased requirements might include the following. (i) A much expanded folk psychology: later hominins routinely engage in joint planning and collective action, and this requires information about the capacities and intentions of social partners. (ii) Later hominins need to understand human symbolic systems, most obviously language, but depictive representation, and signals of age, status and role encoded in dress and action. (iii) A successful agent in recent hominin social worlds needs to understand the norms and normatively laden customs of that world. (iv) Our core group size has probably expanded. So we each will need an expanded database of individual agents: a database that will include partial biography and social assessment, not just recognitional capacity (Dunbar 2001, 2003). (v) Human social worlds have become vertically complex. Human social worlds include teams, extended families, clans, hunting alliances and many other stable, functionally important units intermediate between individual agents and the social world as a whole. Effective action requires the capacity to recognize and work within these intermediate units. (vi) Our technology has expanded explosively, and while (especially in contemporary worlds) no one individual is a master of all the technological resources of his or her group, each agent has the capacity to use (and sometimes make or repair) many tools and their products. Compared with our distant ancestors, we are all engineers. (vii) We have long lived in a world of trade and division of labour. So we must know the economic or exchange value of many goods.

Many of these changes have ancient origins, but they have all intensified (or originated) with the transition to large-scale social worlds. Over the last 10,000 years, human social worlds have become larger and more individually heterogenous. They have more hierarchical structure and more occupational complexity. At roughly the same time, human symbolic systems have become more complex with the invention of writing, the elaboration of numeracy supported by increasingly powerful notational systems and with the expansion of depictive representation. Human social worlds have become vertically complex, and more technologically and economically complex. We need more informational resources to make adaptive choices.

Fortunately for our capacity to achieve our goals, our access to information has improved too. Human learning environments are adapted, making trial-and-error learning more effective, and our technology includes informational technology: most obviously language. Moreover, humans invest time in explicit teaching. So despite the expansion of our informational needs, the shift to large-scale societies may not have exacerbated the problem of information stress. The transition to mass society increases informational demands on adaptive choice, but it also provides more information tools. In key domains of action, we have continued access to instrumentally relevant information, and a continued ability to use that information. This explains the viability of rational agent models of human action. In contrast to the models of behavioural ecology, they can be neutral about the connection between utility and fitness.

11.4.3 Wanting what our genes need

The course of hominin evolution has transformed not just hominin access to information; it has also transformed the motivational structure of human minds, including making our motivations much more sensitive to our social environment. Our ancestors were not chimps. But I shall follow others in treating chimps and bonobos as rough analogues of the minds and capacities of early hominins. Chimp goals are very different from those of modern humans (Tomasello 2009). Some of the most salient differences are:

(i) Humans are not always good at deferring gratification. But we routinely engage in planned activity whose rewards are hours, days, weeks in the future. Many of us save, deferring reward for years and decades. (We do so, in part, by structuring our environment to make the temptations to take immediate reward less available or less tempting.) Chimps have very high discount rates: they find it very difficult to defer reward.[6]

(ii) Chimps are less strongly motivated by social emotions than are humans. They will engage in some low-cost prosocial helping

[6] Though see Rosati et al. 2007; Osvath and Osvath 2008.

(Warneken and Tomasello 2006; Warneken et al. 2007), but there is no sign that they are motivated by such moral emotions as fairness or a preference for egalitarian outcomes. Thus, unlike humans, they seem to act like economic maximizers in ultimatum games (Jensen et al. 2007). In strong contrast to young children, there is no evidence that young chimps find collective activity intrinsically rewarding (Warneken, forthcoming).

(iii) There is no sign that chimp social interaction is regulated by norms that the agents themselves have internalized.

(iv) As with any animal, we feel hungry and thirsty. But the foods we eat, and the circumstances in which we eat, have been much modified by culture, and in ways that vary significantly from culture to culture. Disgusts and taboos show significant cultural variation. In many cultures, resource consumption has acquired an additional signalling function, so our consumptive appetites are mixed with social motivations. Food isn't just fuel; for us, basic biological needs have been infected with cultural significance (Jones 2007). One of the charms of watching chimps at the zoo is their apparent freedom from anything that corresponds to the human emotion of embarrassment, especially in conjunction with lust, hunger or defecation.

(v) We want what others want, because they want it. In part, no doubt, this is a decent but fallible epistemic heuristic: if a book is a best-seller, and I have no reason to regard my own tastes as unusual, I am quite likely to enjoy it too. But as Frank argues (Frank 2000), our interest in what others have is not just instrumental. How I value what I get depends in part on what you get, and on your responses to what I get.

(vi) In contrast to other primates, we have second-order preferences. I can be strongly motivated by the desire for chocolate, but also desire not to want chocolate. While second-order preferences do not inevitably trump first-order ones, and most certainly do not extinguish them, they are not epiphenomenal. In the short-run, second-order desires can block or modify acting on first-order desires, and in the long-run, changes in habitual patterns of action change first-order motivation. I gave up sugar in tea and coffee for health reasons, but no longer have any desire for sweetened caffeine. Mark Hauser discusses more dramatic cases of this interaction, in discussing moral vegetarianism. Such vegetarians often initially find it difficult to give up meat eating, but over time come to find the prospect of meat eating disgusting (Hauser 2006).

As a result of these evolved changes in our motivational structure, our motivations are, at least in part, malleable over an individual's life, as well as on multigeneration scales. To some extent, we learn what we want, not just how to get what we want. Sometimes this is deliberate, as people engineer their first-order desires by changing their way of life to suppress temptation and to enhance the attractiveness of other options. But sometimes there are gradual, unnoticed changes that are side effects of change in the larger environment. Cultural processes install taste and taboo as children mature in their society, but these are not always fixed immutably. The western world has seen major changes in food, dress and sexual norms in the last few decades, and while these changes are reflected most obviously in those who grow up with the new customs, some members of older generations change (perhaps partially) with the times. We are not stuck with the proximate motivators that drove our behaviour at twenty, and not just because of background changes in our intrinsic physiology.

It follows from this plasticity of conscious desire that the targets of individual action are (perhaps profoundly) influenced by variable cultural factors. With the erosion of heritability of phenotypic traits that depend on cultural transmission, and decreased power of cultural group selection, population-level processes are less effective in binding proximal motivation to fitness effects. They are less effective in linking human wants to genetic needs. We can monitor and improve our performance as rational agents in ways that do not depend on population-level processes. But this will not weld human interests to genetic interests. If we get it wrong about what the world is like, often we get a useful though unpleasant signal from the world, and we can use that signal to update and improve our image of the world. An agent can use feedback from the world both to improve his/her stock of instrumental information and to fine-tune techniques for extracting information from the environment. If an agent's proximate motivators come to covary less well with fitness – if the targets at which action is aimed become irrelevant or detrimental to fitness – there need be no consequence in the lifespan of the agent that the agent can recognize and use. Indeed, from the agent's perspective, nothing has gone wrong. A mismatch between what agents value and resources for fitness will not cause the agent's proximate projects to miscarry. So there are no mechanisms operating within the span of individual learning which will systematically cause individuals to unlearn maladaptive values.[7]

11.4.4 *Upshot*

In traditional social worlds, dominated by vertical information flow and group selection, proximate projects guided by maladaptive values will have population-level consequences. Agents with such values will decline in frequency over time. But if vertical information and value flow is less salient and group selection is insignificant, there are no population-level mechanisms which filter maladaptive values (except perhaps in extreme cases). In brief: the transition to mass society need not undermine instrumental rationality; in favourable cases, we can estimate the outcomes of action. But the targets at which we aim need not contribute systematically to fitness maximization. Hence when mechanisms of instrumental assessment are still intact, some form of a utility-maximization model is likely to capture central phenomena of mass societies. In key cases, agents give themselves their best chance of getting what they want; but what they want is no longer reliably a fitness resource. Utility is decoupled from fitness.

REFERENCES

Binmore, K. (2006). 'Why Do People Cooperate?' *Politics, Philosophy and Economics* 5: 81–96.

Bliege Bird, R., E. A. Smith and D. W. Bird (2001). 'The Hunting Handicap: Costly Signaling in Human Foraging Strategies'. *Behavioral Ecology and Sociobiology* 50(1): 9–19.

Blurton Jones, N. (1997a). 'Bushman Birth Spacing: A Test for Optimal Interbirth Intervals'. *Human Nature*, ed. L. Betzig. Oxford, Oxford University Press, 73–82.

(1997b). 'Too Good To Be True? Is There Really a Trade-Off between Number and Care in Human Reproduction?" *Human Nature*, ed. L. Betzig. Oxford University Press, 83–86.

Blurton Jones, N., K. Hawkes and J. F. O'Connell. (2006). 'The Global Process and Local Ecology: How Should We Explain Differences between the Hadza and the !Kung?' *Cultural Diversity among Twentieth Century Foragers: An African Perspective*, ed. S. Kent. Cambridge University Press, 159–187.

[7] Thus in contrast to Don Ross, I do not think that in the long run, fitness must count, and that culturally supported but fitness-eroding values and norms will be filtered out by population-level mechanisms (Ross 2006, 2008).

Boehm, C. (1999). *Hierarchy in the Forest*. Cambridge, MA, Harvard University Press.

Bowles, S. and H. Gintis (2006). 'The Evolutionary Basis of Collective Action'. *The Oxford Handbook of Political Economy*, ed. B. Weingast and D. Wittman. Oxford University Press, 951–970.

Boyd, R., H. Gintis, S. Bowles and P. J. Richerson (2005). 'The Evolution of Altruistic Punishment'. *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*, ed. H. Gintis, S. Bowles, R. Boyd and E. Fehr. Cambridge, MIT Press, 215–227.

Boyd, R. and P. Richerson (1992). 'Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups'. *Ethology and Sociobiology* 13: 171–195.

Crosby, A. W. (2004). *Ecological Imperialism: The Biological Expansion of Europe*, 900–1900. Cambridge University Press.

Diamond, J. (1998). *Guns, Germs and Steel: The Fates of Human Societies*. New York, W.W. Norton.

Dunbar, R. (2001). 'Brains on Two Legs: Group Size and the Evolution of Intelligence'. *Tree of Origin*, ed. F. de Waal. Cambridge, MA, Harvard University Press, 173–192.

(2003). 'The Social Brain: Mind, Language and Society in Evolutionary Perspective'. *Annual Review of Anthropology* 32: 163–181.

Fehr, E. and U. Fischbacher (2003). 'The Nature of Human Altruism'. *Nature* 425: 785–791.

Fehr, E. and S. Gächter (2002). 'Altruistic Punishment in Humans'. *Nature* 415(10 January): 137–140.

Frank, R. (1988). *Passion within Reason: The Strategic Role of the Emotions*. New York, W. W. Norton.

(2000). *Luxury Fever: Money and Happiness in an Era of Excess*. Princeton University Press.

Gächter, S. and B. Herrmann (2009). 'Reciprocity, Culture and Human Cooperation: Previous Insights and a New Cross-Cultural Experiment'. *Philosophical Transactions Royal Society B* 364(1518): 791–806.

Gintis, H. (2006). 'Behavioral Ethics Meets Natural Justice'. *Politics, Philosophy and Economics* 5(1): 5–32.

(2007). 'A Framework for the Unification of the Behavioral Sciences'. *Behavioral and Brain Sciences* 30(1): 1–61.

(2009). *The Bounds of Reason: Game Theory and the Unification of the Behavioural Sciences.* Princeton University Press.

Godfrey-Smith, P. (2006). 'The Strategy of Model-Based Science'. *Biology and Philosophy* 21: 725–740.

(2009). 'Models and Fictions in Science'. *Philosophical Studies* 143: 101–116.

Gurven, M. and K. Hill (2009). 'Why Do Men Hunt? A Reevaluation of "Man the Hunter" and the Sexual Division of Labor'. *Current Anthropology* 50(1): 51–74.

Hauser, M. (2006). *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York, HarperCollins.

Hawkes, K. and R. Bird (2002). 'Showing Off, Handicap Signaling and the Evolution of Men's Work'. *Evolutionary Anthropology* 11(1): 58–67.
Henrich, J., R. Boyd, S. Bowles, et al. (2005). '"Economic Man" in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies'. *Behavioral and Brain Sciences* 28: 795–855.
Hrdy, S. B. (2009). *Mothers and Others*. Cambridge, MA, Harvard University Press.
Jensen, K., J. Call, and M. Tomasello (2007). 'Chimpanzees Are Rational Maximizers in an Ultimatum Game'. *Science* 318(5 October): 107–109.
Jones, M. (2007). *Feast: Why Humans Share Food*. Oxford University Press.
Jukema, J. and T. Piersma (2006). 'Permanent Female Mimics in a Lekking Shorebird'. *Biology Letters* 2(2): 161–164.
Kaplan, H. (1996). 'A Theory of Fertility and Parental Investment in Traditional and Modern Human Societies'. *American Journal of Physical Anthropology* 101(S23): 91–135.
Keen, I. (2006). 'Constraints on the Development of Enduring Inequalities in Late Holocene Australia'. *Current Anthropology* 47(1): 7–38.
Klein, R. and B. Edgar (2002). *The Dawn of Human Culture*. New York, Wiley.
Krebs, J. and N. Davies, eds. (1997). *Behavioural Ecology: An Evolutionary Approach*. Oxford, Blackwell.
Laland, K. (2007). 'Niche Construction, Human Behavioural Ecology and Evolutionary Psychology'. *Oxford Handbook of Evolutionary Psychology*, ed. R. Dunbar and L. Barrett. Oxford University Press, 35–48.
Laland, K. and G. Brown (2002). *Sense and Nonsense: Evolutionary Perspectives on Human Behaviour*. Oxford University Press.
Marlowe, F. W. (2000). 'Paternal Investment and the Human Mating System'. *Behavioural Processes* 51: 45–61.
Norman, M., J. Finn and T. Tregenza (1999). 'Female Impersonation as an Alternative Reproductive Strategy in Giant Cuttlefish'. *Proceedings of the Royal Society B* 266: 1347–1349.
Osvath, M. and H. Osvath (2008). 'Chimpanzee (*Pan troglodytes*) and Orangutan (*Pongo abelii*) Forethought: Self-Control and Pre-experience in the Face of Future Tool Use'. *Animal Cognition* 11(4): 661–674.
Quinn, T. (2005). *The Behavior and Ecology of Pacific Salmon and Trout.* Seattle, University of Washington Press.
Richerson, P. and R. Boyd (2001). 'The Evolution of Subjective Commitment to Groups: A Tribal Instincts Hypothesis'. *Evolution and the Capacity for Commitment*, ed. R. Nesse. New York, Russell Sage Foundation, 186–220.
Richerson, P., R. Boyd and J. Henrich (2003). 'Cultural Evolution of Human Co-operation'. *Genetic and Cultural Evolution of Cooperation*, ed. P. Hammerstein. Cambridge, MIT Press, 373–404.
Rosati, A., J. Stevens, B. Hare and M. D. Hauser (2007). 'The Evolutionary Origins of Human Patience: Temporal Preferences in Chimpanzees, Bonobos, and Human Adults'. *Current Biology* 17: 1663–1668.
Ross, D. (2006a). 'Evolutionary Game Theory and the Normative Theory of Institutional Design: Binmore and Behavioral Economics'. *Politics, Philosophy and Economics* 5(1): 51–80.

(2006b). 'The Economic and Evolutionary Basis of Selves'. *Cognitive Systems Research* 7: 246–258.

(2008). 'Economics, Cognitive Science and Social Cognition'. *Cognitive Systems Research* 9: 125–135.

Seabright, P. (2006). 'The Evolution of Fairness Norms: an essay on Ken Binmore's Natural Justice'. *Politics, Philosophy and Economics* 5(1): 33–50.

(2010). *The Company of Strangers*. Princeton University Press.

Smith, E. A., K. Hill, F. Marlowe, et al. (2010). 'Wealth Transmission and Inequality among Hunter-Gatherers'. *Current Anthropology* 51(1): 19–34.

Smith, E. A. and B. Winterhalder (2003). 'Human Behavioral Ecology'. *Encyclopedia of Cognitive Science*. L. Nadel. London, Nature Publishing Group. 2: 377–385.

Sober, E. and D. S. Wilson (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA, Harvard University Press.

Sterelny, K. (2007). 'Social Intelligence, Human Intelligence and Niche Construction'. *Proceedings of the Royal Society B* 362(1480): 719–730.

(forthcoming). 'Life in Interesting Times: Co-operation and Collective Action in the Holocene'. *Evolution, Cooperation, and Complexity*, ed. B. Calcott, B. Fraser, R. Joyce and K. Sterelny. Cambridge, MA, MIT Press.

Tomasello, M. (2009). *Why We Cooperate*. Cambridge, MA, MIT Press.

Warneken, F. (forthcoming). 'The Origins of Human Cooperation from a Developmental and Comparative Perspective'. *The Evolution of Mind*, ed. G. Hatfield. Philadelphia, University of Pennsylvania Press.

Warneken, F., B. Hare, A. P. Melis, D. Hanus and M. Tomasello (2007). 'Spontaneous Altruism by Chimpanzees and Young Children'. *PLoS Biology* 5(7): e184, doi:10.1371/journal.pbio.0050184.

Warneken, F. and M. Tomasello (2006). 'Altruistic Helping in Human Infants and Young Chimpanzees'. *Science* 311(3 March): 1301–1303.

Weisberg, M. (2007). 'Who Is a Modeler?' *British Journal for the Philosophy of Science* 75(3): 383–404.

Wimsatt, W. C. (2007). *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA, Harvard University Press.

Winterhalder, B. and E. A. Smith (2000). 'Analyzing Adaptive Strategies: Human Behavioral Ecology at Twenty Five'. *Evolutionary Anthropology* 9(2): 51–72.

Zahavi, A. (1990). 'Arabian Babblers: The Quest for Social Status in a Cooperative Breeder'. *Cooperative Breeding in Birds*, ed. P. Stacey and W. Koenig. Cambridge University Press, 105–133.

Zizsser, H. (1935). *Rats, Lice, and History*. London, George Routledge & Sons.