

**My Nipples Explode with Delight:
Signals, Co-operation, Learning**

**Paper for Jean Nicod Lectures, 2008
Background to Lecture 3**

**Kim Sterelny
Philosophy and Tempo and Mode
VUW & ANU**

May 2008

1. Sperber's Dilemma

We are habitual and obligate participants in rich networks of social or cultural learning. Much of the information (and misinformation) in my head is there because it was, earlier, in the heads of other agents. I am an information soak. I am also an information source: information (and misinformation) is in many other heads because it was first in mine. While social learning is far from being uniquely human, its nature and role has been transformed in our lineage (Tomasello 1999; Richerson and Boyd 2005). In many species, agents learn from other adults about immediate threats and opportunities (as, for example, when birds in a flock respond to others' alarm calls). Indeed, such social learning often does not depend on others signalling at all; one animal learns from another through information leakage. One bird in a flock spooks at a hawk overhead, and the surrounding birds notice and respond to its reaction. They benefit from one another's response, but the initial response is independent of the benefit it generates¹. In contrast, much human cultural learning depends on signalling. Moreover, our cultural learning is often inter-generational: our signals transmit long-life information, and often using arbitrary, low-cost signals (Laland and Hoppitt 2003; Danchin and Luc-Alain 2004).

Social learning, then, has a distinctive role in our lifeways. Moreover, we also habitually think and communicate about thinking and communication. Uniquely amongst the primates, we fluently metarepresent. In an insightful series of papers, Dan Sperber has connected these two features of human cognitive life, arguing that the evolution of fluent metarepresentation is more directly connected to the elaboration of agent-to-agent communication than to the elaboration of individual cognitive capacity. Metarepresentational capacities, he suggests, evolved largely as a response to the increased importance of signalling — of cultural learning — in human life (Sperber 1997; Sperber 2000; Sperber 2001).

¹ This benefit of social life is known as “by-product mutualism”, in the trade.

Sperber begins with the crucial idea that communication and cultural learning is a special case of co-operation². As with other forms of co-operation, communication (especially communication using cheap and arbitrary signals) is both risky and potentially profitable. Listening to another agent communicate about their common world seems to offer the opportunity to acquire at negligible cost crucial information. Agents could often discover this information for themselves only with real effort or risk; sometimes individual discovery will not be possible at all. Moreover information can be of great value. Information about threats and opportunities can determine the course of one's life. So the potential benefit is extremely high; equally, the costs of trusting another can be catastrophic³. There are, for example, few disasters more frequent and less repairable than having children with the wrong partner. In an environment of frequent informational co-operation and communication, the rewards are too great to be forgone. But the risks are too great to be ignored. Yet just in those cases where the benefits are greatest — where communication carries information about aspects of the world that are both important but which are expensive or impossible for the soaker to access — the veracity of the signal cannot be directly checked. So Sperber's dilemma is the dilemma of trust: we cannot afford not to trust, and we cannot afford to trust the faithless.

There are, in fact, two defector-driven threats to co-operative information sharing. The threat of free-riding presupposes that collecting information is not free: a free-riding agent does not collect information, and thus shares in the benefits of pooled information while paying none of the costs. Deception involves sending signals that alter the behaviour of the receiver in ways that have fitness costs to the receiver and benefits to the sender. The evolution of information sharing via arbitrary signals creates an opportunity to exploit an information soaker via manipulation through false signals. Sperber's problem of trust is focused mostly on this second problem, and while there is no perfect solution, Sperber suggests that folk epistemology — our set

² Of course, Sperber is not the only one to see this connection, though his presentation is particularly vivid: the same general problem is also explored in (Cosmides and Tooby 2000; Bacharach and Gambetta 2001; Gambetta 2005).

³ On this way of modelling information-sharing, the payoff structure is quite different from that used to model the evolution of such conditionally co-operative strategies as tit-for-tat: tit-for-tat succeeds in a mixed environment only because the sucker's payoff was not crippling (and a single payout from successful defection was not enriching).

of tools for representing and evaluating signs and signals — is a partial solution. This folk epistemology is most naturally seen as a device which helps stabilise informational co-operation once it has evolved, thus creating a niche for deception. Metarepresentation evolves as part of a mechanism of indirect scrutiny; as part of a folk epistemology. We have a folk logic. We do not just represent representations; we assess them⁴. These capacities enable us to assess messages for their coherence with what we know from other sources, and with what the agent has previously said. And we can keep track of a source in order to build an epistemic profile of that source: we assess the reliability of sources as well as the plausibility of messages. These precautions are not perfect, and they are not free. But they are part of the trade-offs involved in trying to maximise the benefits of informational trafficking while minimising the risks. Sperber’s model is a stabilisation model: it is not intended to explain how or why elaborate cultural learning evolved: it explains how such learning can persist in the face of pressures to decay. While developing a different perspective, the focus of this paper, too, is on the stabilisation, not the construction, of our elaborate machine of cultural learning.

In short, Sperber proposed an antiviral model of folk epistemology. To the extent he is right, we expect the sophistication of folk epistemology to track the threat of free riding and deception. So, since low-cost arbitrary signals are more exploitable than channels which are intrinsically more reliable (acquiring a skill by observation learning, for example), the folk epistemology of such risky channels should be better developed. There is some hints that this is right. Think of language, the paradigm arbitrary, low-cost channel. “Nonsense”, “bullshit”, “doesn’t follow” are all sceptical assessments of linguistic signals; these and other folk concepts are part of our standard operating equipment. Contrast leakage. It is possible to (for example) fake the signs of age and declining health. No doubt professional actors have a specialist vocabulary to identify the relevant parameters of gait, stance and movement through which signs of physical condition leak. But these are not standard items of folk equipment. For this is not a channel through which our model of the environment is

⁴ Since referential signals by epistemically imperfect agents in a noisy and confusing world can never have been perfectly reliable, regular referential communication must have always required assessments of reliability; even between completely honest agents, blind trust would not have been a good possibility. So folk logic presumably evolved out of assessing the reliability of referential signals.

routinely manipulated. Just as we should be more wary about some channels than others, we should also be more wary in some social worlds than others. Thus we should expect folk epistemology to be better developed in cultures in which deception and manipulation are more of a threat: presumably cultures which are larger; more socially structured; in which agents have more one-off interactions, and in which agents are more mobile. Contemporary mass societies are at one limit here, so it would be instructive to compare folk epistemology skills cross-culturally, between our large open heterogeneous societies and small, fairly closed, traditional ones. A Sperberian model of the function of folk epistemology predicts that we are good folk epistemologists, for we live in social worlds of unprecedentedly high threat⁵.

I think there is something importantly right about this idea, but this perspective on folk epistemology and its role in filtering deceptive signals is too Machiavellian. The problem of trust is genuine and ancient: free-riding and deceptive manipulation is a risk to those engaging in information sharing. This threat is clearly much more serious in contemporary mass social worlds, with their many one-off and arms-length interactions. It is hard to think of a plausible Pleistocene analog of Nigerian e-mail scams. But manipulation is a threat even in the small scale, intimate social worlds of most human evolutionary history. But though real, this threat is not uniform. I shall suggest that many information-sharing interactions are not seriously threatened by either free-riding or by deceptive manipulations, even ones exploiting low-cost arbitrary signals. So I agree that Sperber's dilemma is important. But I do not think it is ubiquitous. Moreover, policing defection is not the only role folk epistemology plays in the evolutionary elaboration of information storing. Like Sperber, I see folk epistemology as a communication tool. But I see it as a multi-purpose tool. One purpose is to detect cheats. But another is to optimise the flow of honest communication. We can and do use broadly metarepresentational capacities to organise and improve the bandwidth and fidelity of information sharing; it is a folk education tool, too. I shall suggest that the set of cognitive capacities Sperber identifies plays a broader role in the organization and optimisation of cultural learning; a broader role that is as ancient as the policing function which he has identified.

⁵ Presuming, of course, that folk epistemology is to some significant degree developmentally plastic.

Moreover, while I agree that these capacities are communication tools, they also play an egocentric role. Once the capacity to represent signals and thoughts is in place, agents can and do represent their own signals and thoughts, and doing so is an important cognitive aid to self improvement. By representing both our own cognitive products, and our route to those products, we can improve our cognitive routines. The limit of this improvement via representation is found in the natural sciences: every branch of the natural sciences has specialists who calibrate and improve the tools of their discipline. They improve: statistical and computational methods; experimental design; measuring instruments; model systems. But there are many humdrum examples of this general phenomenon. I am monocular, and know it, and as a consequence know that my depth vision and judgement of distance, especially at night, are not reliable. I have to be more cautious than others in crossing roads at night. For the same reason, I have never acquired a driver's license.

So while metarepresentational capacities are used in making cultural learning more efficient, they are also used in making individual learning and decision more efficient. So, for example, to use a spear as a template for spear-making, an agent must understand that a spear is an information source, not just a weapon. Metarepresentational capacities are used to access and use cognitive tools, and while those are often used in interpersonal contexts, they are not only used in those contexts. Andy Clark argues that these "second order cognitive dynamics" (as he calls them) are powerful and important: we can recognise flaws in our plans, judgements or arguments, and devote cognitive resources to fixing such plans; we can become aware of the unreliability of our prima facie judgements about certain phenomena and in the short term proceed with special caution while longer term devote resources to improving our own reliability; we can recognise the circumstances in which our judgements are most reliable, and try to ensure that we make our most important decisions in those circumstances. Thinking about our own thinking is far from epiphenomenal ((Clark forthcoming), 3.6). Agents can and do represent their own cognitive products, and their doing so has consequences for their own cognition and behaviour.

2. Two Faces of Cultural Learning: *Diplomacy*TM and Monty Python's Hungarian Phrase Book.

Sperber's dilemma is real, but not ubiquitous: foul dealing is a threat to some information sharing transactions but not others. For the profit of cultural learning and its potential risks vary: they are contingent on the identities of source and soak; on the domain about which communication takes place; and on the communication channel. Before attempting a preliminary analysis of these complications, I shall begin with a couple of illustrative examples. The first exemplifies the full-on Machiavellian dynamics that Sperber's analysis tracks. The second is a contrast case.

In the days before computer games took over, *Diplomacy* was a popular, though relationship-stressing board game. The object of the game was to build a Europe-dominating empire, through a judicious combination of alliance and betrayal. In *Diplomacy*, the paradigm communicative act was the conspiratorial whisper; the paradigm topic of conversation concerned intentions to future actions. Bracketing off the fact that costs and rewards in *Diplomacy* were mostly pretence, the communicative situation fits the Sperberian paradigm perfectly. The dynamic is Machiavellian. Agents are self-interested, but there is no triumph without alliance; no alliance without the risk of betrayal. Information soaks have no independent, direct test of signal veracity (until it is too late). But since blind trust is fatal, imperfect indirect tests must be used. In particular, a soak must decide whether a source's stated intentions cohere with those that the soak identifies as the source's optional choices, given the soak's assessment of the situation the source takes himself to be in. Folk epistemology is a fallible tool, but it is the best agents have in managing and assessing conspiratorial whispers.

Diplomacy is indeed a model of one form of cultural learning and communication, and in such cases folk epistemology does indeed play much the role Sperber identifies. But *Diplomacy* is not the only model of cultural learning and communication. Consider a contrasting example from my youth: Monty Python's celebrated Hungarian Phrase Book sketch, in which a publisher produces a deceptive supposedly English-Hungarian-English phrase book in which, for example, the

Hungarian phrase meaning "Can you direct me to the station?" is translated by the English phrase, "Please fondle my bum" and a protest about false arrest becomes "my nipples explode with delight". To those ignorant of English (or Hungarian), the adequacy of this translation is difficult to check directly (again, as in the first example, until it is too late). But there could very rarely be a temptation to deceive in such a case. In part, this is because the phrase-book is a public broadcast, rather than a signal to a specific, pre-identified agent (hence the consequences of successful deception are much less easy to identify). For the same reason, successful deception is much harder to manage: not all those who receive a widely broadcast signal will be ignorant in ways that make them vulnerable to manipulation. Their response can then cue those who are ignorant. The channel (language) is the same as in *Diplomacy* conspiracies, but with the change in topic, and with the change to a multi-agent, epistemically heterogeneous audience, the threat of deception essentially vanishes. Not all phrase books are well-designed, but we discount the problem of manipulation for good reason.

These are toy examples, but they begin to reveal the complexity of cultural learning, and the variability of the threat of deception. In the discussion to come, I make four central points. (i) As Sperber and others insist, information sharing is indeed an instance of co-operation, and like other forms of co-operation, there are often possibilities of free riding and deception. But this threat is not uniform. Source and sink; domain; the potential benefits of information sharing; signal channel vary. Those variations are all relevant to the existence and severity of defection and deception problems.

(ii) In part because the threat of defection is so variable, I shall suggest that folk epistemology is not just a policing mechanism. It is not just a filter that suppresses deception by making its detection more likely. It is also a set of tools we use to enhance the efficiency of agent-to-agent information transfer (and individual exploration, but that will not be my concern here).

(iii) This variability underscores another point that is obvious once made. The evolution of cultural learning and information sharing is not a unitary phenomenon.

Rather: it is a complex of coevolving but somewhat separate capacities. These include gesture and mime; language; theory of mind; observation learning. We have evolved the capacity to transmit and read signals sent through many channels; channels which vary in their reliability, their bandwidth, and in the kind of information that can be sent through them. We have also evolved a range of ways of monitoring and intervening on those channels.

(iv) Finally, the evolution of the cultural learning complex is not just the evolution of such individual adaptations. Cultural learning is an interaction effect. Crucial forms of cultural learning depend on the construction and stabilization of social worlds of the right kind. This might seem trivial: whatever their individual cognitive capacities, two agents cannot share information unless they are in contact. But it is not trivial: much more than mere contact is needed for the extensive and essential information sharing that is central to human existence.

3. The Cultural Learning Complex.

I begin by representing the complexity of human cultural learning through two tables. The first identifies a set of content domains; identifies the typical relationship between agents that share such information, and the typical costs and benefits of sharing to the relevant parties. The second identifies a set of channels; assesses the intrinsic reliability of these channels, and specifies the type of information that typically flows through them. I do not want to claim too much for these tables. They are not complete, and nor do they represent the only way of segmenting the cultural learning complex into components. Moreover, there are many hybrids: we learn about others' reliability, for example, by what they say but also what they do (information leaks as a result of their non-communicative activity); from what third parties say about them, but also from their body language, their tone of voice and expression in interactions, and by third party economic interactions.

Still, despite the limitations of this analysis, it does highlight the complexity of human cultural learning, and the subtle interplays between cost, benefit and reliability; an interplay that will depend on (i) sources, soaks and their relationship; (ii) the domain:

the type of information the soak pumps from the source; (iii) the information channel(s) — the mode or modes through which information flows. The tables do show that the elements which constitute the human cultural learning syndrome vary in (a) fidelity and bandwidth: the noise that degrades information, and the amount of information that flows; (b) potential temptations to deceive; (c) the intrinsic reliability of signal and cue types and the extent to which the soak can independently assess the veracity of a signal. The analysis also shows that the standard model of the evolution cultural learning is brutally over-simplified. In modelling the evolution of cultural learning, that standard model takes cultural learning to be cheaper (because it avoids costs of time and error) but less accurate than individual, trial and error learning. Saving those learning costs selects for extracting information from others, so long as the environment does not change so fast that others' information is not hopelessly put of date (see for example (Boyd and Richerson 1996; Laland 2001). Information soaks are conceptualised as information parasites, somewhat degrading the overall reliability of the informational resources of the group. As we shall see, this is a very misleading model of cultural learning and its evolution.

Content Domain	Source benefit/loss	Soak benefit/loss	Collective benefit	Source/Soak information asymmetry
<p>Long-life information re local environment.</p> <p>It is possible, but sometimes expensive to acquire this information by individual learning.</p>	<p>Apparent relative fitness sacrifice</p>	<p>High benefits; crucial information.</p>	<p>Collective manages local environment more efficiently; buffering group info resources</p>	<p>Information often flows both ways; often many/many interactions</p>
<p>short-life information re local environment.</p> <p>It is usually possible but sometimes expensive to acquire this information non-socially. Time constraints might make it impossible</p>	<p>Apparent relative fitness sacrifice; but information pooling can increase reliability</p>	<p>Variable. Sometimes crucial information</p>	<p>Threat/opportunity management; information pooling increases reliability</p>	<p>Information often flows both ways; often many/many interactions</p>
<p>Skills</p> <p>Often not possible to acquire skills by individual learning</p>	<p>Apparent relative fitness sacrifice; but between adults sometimes gain by increase in reliability of</p>	<p>Very high benefits; crucial life skills not otherwise to be had</p>	<p>Division of labour; benefits of specialisation; buffering group info resources; extending platform for further innovation</p>	<p>Typically asymmetrical</p>

	information pooling			
Third-party social information (gossip). Sometimes possible but expensive to acquire by individual learning	Source may benefit via defection control	Choosing right social partners; high risk-high reward decisions	Control of defection; policing norms	Information often flows both ways; often many/many interactions
Second-party social information (boasting) Sometimes possible but expensive to acquire by individual learning	Source may benefit via beneficial relation with soak	Choosing social partners; high risk-high reward decisions	neutral	Usually unidirectional
Co-ordination information/bargaining information. Typically this information must be acquired culturally	Managing mutual exchange; Source may benefit via beneficial relation with soak	Managing division of labour & reciprocation-based co-operation	Division of labour; collective decision making?	Information flows both ways
Local customs, mores, norms Typically this information must be acquired culturally	Source may benefit via social co-ordination	Avoiding punishment; smoother co-ordination with social partners	Variable, as customs/norms may not be adaptive	Usually unidirectional

<p>Public representational media</p> <p>Typically this information must be acquired culturally</p>	<p>Source gains new tools for influencing soak's behaviour</p>	<p>Soak gains cognitive & communication resources</p>	<p>Division of labour; improved co-ordination; buffering group info resources</p>	<p>Usually unidirectional</p>
--	--	---	---	-------------------------------

Channel	Intrinsic Reliability	Domain
Guided/structured trial and error	high	Skills; long-life information re local environment
Imitation learning via demonstrations	high	skills
Mimesis, gesture, depictive representation	medium	Domain general, but perhaps not customs and norms
language	low	Domain general, but often supplements rather than replaces other channels
Information leakage via cues and economic activity	Variable (as leaks can be faked/suppressed), but often high	Skills; social information; local ecological information
Costly signals	high	Second-party social information; sometimes bargaining contexts

I do not intend to work through these tables row by row. Rather, I shall use them to make some more general points about the pervasiveness of Sperber's Dilemma and the role of folk epistemology; about the many mechanisms that limit the danger of defection to information sharing, and the light this analysis throws on the evolution of information sharing.

Sharing Ecological Information

Consider, first, the complex of issues around shared information long-shelf-life information about the local environment. The standard model does not capture these dynamics well. In general, an individual can find for themselves some information about the location of resources and dangers; threats and opportunities. But in a

heterogeneous and changing environment, no individual by themselves can find out all the information of this kind that is potentially relevant. Heterogeneity creates an information gradient (Dennett 1983; Dennett 1988), and, hence a potentially advantageous division of epistemic labour. Everything has been seen by someone, but no-one has seen everything; certainly, no-one has seen everything recently. This form of information pooling is not a cheaper but less accurate route to cognitive resources an individual could amass for himself; rather, it makes both individuals (and collectives) more effective epistemic agents. Yet such information sharing seems to be evolutionarily fragile. For the information channels, while varied, are not intrinsically reliable: the information flow depends largely on language, gesture and similarly low-cost signals about spatiotemporally displaced targets of inquiry (though no doubt some information flows inadvertently, via leakage⁶). Since the information flow concerns the elsewhere and the elsewhere, their veracity cannot be checked directly, against the world. And if temptations to deceive exist, the characteristics of the channel itself will not prevent succumbing to those temptations. If the profit of cooperation is generated by reciprocation over time, yet on particular occasions of information donation there is a temptation to defect, the problem of trust for the information soak seems serious. The problem seems especially serious if the signals conveying information are not intrinsically reliable. How do agents in information exchange police fair reciprocation?

Despite the problem of low intrinsic reliability, I do not think Sperber's Dilemma is especially pressing in this class of cases. It is mitigated by two crucial, and I suspect stable and widespread features of human social environments. The first is that information flow is often many-to-many: this makes both free riding and deceptive manipulation much harder (as those in the know will be aware of such defecting behaviour), and the audience, too, will vary in the extent of their ignorance. What might deceive or manipulate one, will not work on another. Public signalling reduces opportunities and temptations to defect. Perhaps even more importantly, in paradigm cases the agents are symmetrical: they each have a chunk of the total informational resources of the group, and none know in advance whose chunk is the most important.

⁶ For example, the resources someone brings back from a collecting trip will tell onlookers something about the conditions in the places she has been. Not coming back at all might hint at unanticipated dangers.

This symmetry means that direct reciprocation is a plausible mechanism that might select for ecological information pooling. Moreover, because this information has a long useful life, information can be and often is pooled prior to individual and collective deliberation and action. There is less temptation to defect because agents often will not be able to assess the value of their private fraction of local knowledge. If people typically pump information into a common pot, there is less temptation to manipulate, because an agent planting false trails will often not know who will act on them or how. Sharing information in advance of action imposes a kind of veil of ignorance that severs the planning connection between false signal and Machiavellian consequences. To phrase the same idea more carefully: to the extent that local knowledge sharing is public and decoupled from immediate action, temptations to defect are eroded. The upshot, then, is that in public signalling contexts, the chance that an attempted manipulation will be detected is quite high, and its rewards will rarely be both high and certain. Since the individual and collective benefits of local knowledge pooling are significant, we can expect a default for honest signalling.

Finally, it may be important that the benefits of ecological information pooling do not depend on ultra-sophisticated communication systems. In small scale, closed communities the importance of reputation — the effects of your acts on third party opinion — is likely to be a powerful mechanism enforcing conformity to norms, including norms of honest communication. But reputation-based enforcement of honesty probably depends on complex communicative capacities (to gossip effectively, I have to be able to represent what others said and did at other times, places and circumstances). Relatively simple protolanguage-style signals, or systems of gesture, mime and depiction, would suffice to signal important environmental information in ways that will be kept honest by public signalling, and by pooling data before action planning. These relatively rudimentary signals can be both honest and cheap; this makes possible the early evolution of environmental information of forms which bring modest benefits to all parties. In contrast, expensive signals can evolve only in high stakes interactions, for the benefit of honest interaction has to pay for the cost of signals.

Redundancy

As we see from the discussion of long-life local information, redundancy plays an important role in suppressing deceptive manipulation. One reason why we need not fear fake Hungarian phrase books is that information about local conventions, customs and norms is typically multiply and repeatedly sourced. Agents rarely learn to read the conventional, low cost signals of their community — language, gesture, “body language”, local marks of status, role, affiliation, group identity — from a single individual; still less, on a single occasion from a single other individual. In my politically depraved youth, I learned the distinctive patois, gesture, attitudes, rituals and public marks of my local Trotskyist tribe by immersion, not by instruction from a single mentor. It would have taken a persistent and disciplined conspiracy (far beyond their organization talents) to practice a deception upon me. We can rely on shared information about norms, customs, symbols in part because we normally acquire information of this kind redundantly and multiply; in part because there is rarely temptation to defect.

However, redundancy can play a second role: information pooling can increase the reliability of judgement in the face of environmental noise, and I suspect that this mechanism might be quite important in the evolution and stabilisation of shared information about rapidly changing features of the immediate environment. Agents live in epistemically polluted environments, because other agents try to both fake and conceal. The dangerous try to look harmless; the harmless try to look dangerous. As a consequence, perceptual signals of opportunity and danger are often hard to interpret, and making them less ambiguous often has a cost. In noisy worlds, there is selection on agents to track salient aspects of their environment by multiple cues: to listen and smell as well as look; to attend to movement as well as colour and pattern; to attend to the responses of other species (why have those parrots just flown?) (Sterelny 2003). The Concordat Jury Theorem makes vivid the value of this shift to multiple channels in the face of noise: so long as each juror votes independently and has a better than .5 chance of being right, as the size of the jury goes up, the probability of a majority vote being right rises rapidly to near-certainty (List 2004).

The jury can be inside the head of a single agent: one that multi-tracks salient features of their environment. But the theorem also captures why herd animals monitor one another's responses. Agents gain access to reliable information about their

environment if there is mutual knowledge of each agent's assessment of noisy signals, together with trust in consensus. This mechanism may well be an important component of co-ordinated activity: imagine a foraging party trying to decide whether a swollen river is too dangerous to ford; which animal in a pack to target; how to interpret the ambiguous behaviour of a neighbouring group. The channels through which mutual knowledge arises are not intrinsically reliable. But there is no temptation to defect here: by voting honestly and accepting consensus, each agent trades an unreliable assessment of a relevant feature of their world for a much more reliable assessment. In principle, agents could go it alone. But testing the river to see whether it really can be forded safely might well be very expensive indeed. A crucial aspect of such cases is that the profit of co-operation does not derive from serial reciprocation; it is immediate. So information-pooling protects not just against heterogeneity but against noise.

Skill

In the cases I have been discussing, the problems of free-riding and defection are limited by the structure of the communicative environment (redundancy; multi-sourcing information; public communication; early pooling of information) and by payoff structures. Free-riding is not much of a temptation, and in most cases there is little reason to expect net benefits from attempts at deceptive manipulation. The story is different with skill. Skill is central to human life and is a key domain of social learning. A central feature of human evolution has been the evolution of cultural transmission of skill that is accurate enough and reliable enough for skills to have become multi-generation products (Hill and Kaplan 1999; Tomasello 1999; Kaplan, Hill et al. 2000; Gurven, Kaplan et al. 2006). So many human skills, including some which are crucial components of fitness, can only be acquired socially. On the face of it, skill transmission sacrifices a relative advantage. But to the extent that skill acquisition depends on a hybrid of socially structured trial and error learning; demonstrations of actual practice, and cues — observations of skilled practitioners using their skills for their own purposes, and observations of the products of those practices — there is little problem of deceptive communication. These channels have high intrinsic reliability. Lies and faking are not a major problem. If a skilled

practitioner signals, and thus demonstrates their own capacities, or guides practice with examples, there is little opportunity for outright deception.

This does over-simplify: skill transmission does sometimes involve arbitrary media: language and gesture. The skills of foragers, artisans and farmers depend on lore as well as know-how. But lore — hunters’ tales about what is found where, and how to catch it — is often multi-sourced and broadcast publicly, and as I have just noted, these are honest signalling mechanisms. Lore is often part of the narrative resources of a community, and while that does not guarantee its veracity, it proofs it against specific deception of an individual by an individual. “Common knowledge” may not be knowledge. But it is not an individually targeted deception. The challenge is to persuade skilled practitioners to signal (Henrich and Gil-White 2001), and given that they have signalled, optimising the learning opportunities that such signals provide. In most circumstances, those with skills will not signal to manipulate and deceive the less skilled. They will signal, (i) if they are kin, and the success of the task contributes to their own fitness; (ii) if skill-signalling is the result of group-level selection for information-based co-operation; (iii) if there is an information market: those who signal are paid, directly or indirectly, for doing so.

High Stakes Negotiations

The threat of deception is most serious when interactions involve high stakes; when communication is private, and uses arbitrary, low-cost signals. In the small scale societies in which humans evolved, these conditions would sometimes be met: most usually in contexts of gossip, negotiation, sexual and political intrigue. These would sometimes create Sperberian dilemmas: the stakes are too high to make opting out of conversational exchange an good option, but the threat of dishonest signals is serious and pressing. For example, Boehm’s Hierarchy in The Forest shows that forager egalitarianism depends on the active defence of those norms and social practices (Boehm 1999). Would-be chiefs do arise in these societies, and sometimes pose such severe threats that active and dangerous measures must be taken against them. More prosaically, sexual negotiation is often a high stakes activity. So there are high stakes co-ordination and partnership decisions in small scale societies, and hence a serious problem of trust. Trust requires more than confidence one is not being lied to: joint

plans miscarry catastrophically through failures of nerve, weakness of will, succumbing to temptation, and sheer incompetence. But agents do lie manipulatively in gossiping about others; they make promises and give guarantees that they never intend to keep. They try to induce others to act in ways that will benefit them, but will have savage consequences for their targets of persuasion. So in social evaluation and social negotiation, we need to and do evaluate both source and message. The evaluation of the source is much aided by leakage: co-operative foraging and other interactions generate rich mutual knowledge, especially of character. In small communities, people know who has their shit together, and who has not. But, clearly, the resources of folk logic will play an important role in managing information about reputation and similar third party social information, and in forming and managing joint activities and ongoing partnerships.

Likewise, costly signalling theory comes into its own in helping explain the limits on defection in these high stakes cases (Zahavi and Zahavi 1997; Saunders forthcoming). This theory explains the signalling dynamics with the systematic temptation to exaggerate found in sexual advertisement and aggressive bluffing, where signallers will always be under selection to exaggerate how sexy, fit or dangerous they are. Despite that temptation, as a consequence of the differential cost of signals — only the really dangerous can afford to seem really dangerous — self-referential signals can still carry real information. The Zahavian route to honesty is irrelevant to many of the cases discussed earlier, for these concern referential signals about the shared environment. Costly signalling theory applies when the signals are about the agent signalling. For it is then that the differential relative costs — in the best cases, the signal can be afforded by and only by agents that signal honestly — can impose honesty on the signalling system.

As Bacharach and Gambetta point out, we can understand trust in information that leaks from the ordinary, utilitarian actions of an agent in the costly signalling framework (Bacharach and Gambetta 2001; Gambetta 2005). Suppose, for example, that Australasian philosophers of biology have a high reputation for probity in all their dealings (which, of course, they do), and as a consequence attractive young women (for example) believe all their promises. It costs me nothing to present as such an agent in my interactions with others. It is possible for, say, a French student of

Foucault to present as an Australasian philosopher of biology, but it would be difficult, hence expensive to do. For it would require massive investment for them to master the cues of that background. Leaks — when genuine leaks — are often information-rich, and while they sometimes can be faked, doing so is hard. Hence cues⁷ can be recruited as signals which are cheap for the honest but expensive for defectors. So in these high stakes interactions, reputation, leakage (and other high signal cost mechanisms) and the resources of folk epistemology will all play a role in determining whether one agent trusts another.

4. The Folk as Educators

In our species, social learning is not a cheaper or more efficient substitute for individual, trial and error learning. In important cases, the information learned socially could be learned no other way. Consider first, information about norms, customs, and the meaning of arbitrary media. In some admittedly obscure sense, information of this kind is essentially social. Consider, for example, distinctive symbols of group identity: say, the tattoos and regalia that a young man wears to show his membership of the motorcycle gang, The Nomads. In the right context, a particular jacket design — a “patch” — symbolises this membership. The facts that constitute the meaning of this symbol are a matter of obscure debate. Even so, it is widely agreed that they are constituted, somehow, by the thoughts, responses and dispositions of the gang members and those that associate with them. To learn these facts about meaning (whatever, exactly, those facts are) is to extract that same information from others’ heads. It is more than just learning some behavioural facts about men that dress that way. A stranger might discover by individual exploration that individuals so patched are dangerous, violent and mostly to be avoided. But that would not suffice for discovering the meaning of the patch. To do that, they have to come to share some of those attitudes and beliefs; to become part of a circle of mutual knowledge.

⁷ Marc Hauser draws this distinction between information that leaks via side-effects of utilitarian action and signals produced with a communicative function in (Hauser 1996), pointing out that their evolutionary dynamics are quite different, but also that the first can morph into the second.

Skills are not similarly intrinsically social. In principle, they can be learned by individuals exploring their physical and biological habitat, and initial breakthroughs to new technologies, resources, and adaptive zones must originally have been made that way. But by the establishment of “behaviourally modern” sapiens (i.e. for more than 100,000 years), the crucial skills of forager life could not be learned individually (McBrearty and Brooks 2000; Richerson and Boyd 2005). Human resource gathering depended on the previous accumulation of information and skill, and the successful transmission of those resources to the next generation. The distinctive style of behaviourally modern foragers — co-operative, technologically and informationally enhanced foraging — generates rich resources from the environment. But it requires up-front investment to assemble the informational resources on which this foraging style depends. Adolescents must be supported while they acquire the high levels of expertise these lifeways demand. Much of this support is informational: expertise is acquired with the aid and support of the upstream generation (Sterelny 2007).

As I have argued before, the acquisition of expertise depends on hybrid learning (Sterelny 2006). No-one learns how to make the key artefacts of traditional societies just by looking and listening. But no-one learns them without looking or listening; skilled performance produces both reference points and products that the less adept can use to guide their practice, even without explicit teaching or staged demonstrations. Such trial and error learning is socially supported in a variety of ways, whose importance will vary from culture to culture and from expertise to expertise⁸. One important mechanism of support is supervised practice. For example, children join adult expeditions, and help adults in tool making and repair under conditions which adults regard as suitable and safe, and which are sometimes modified to make the experience safer or more educationally effective for the young. On such expeditions: the learning is less hazardous, less expensive, and better targeted. Social signals of failure substitute for the world’s signals. Adults intervene in the learning world of children in many other ways. Thus they provide children with toys (for example, the Hazda provide miniature bows; the Merriam, fishing

⁸ There is considerable scepticism about whether social learning is indeed high fidelity; so much so that Robert Boyd and Peter Richerson devote some effort to showing that evolutionary models of culture do not depend on high fidelity transmission (Richerson and Boyd 2005). But these sceptical papers (Atran 2001; Gil-White 2005) do not focus on skill, nor on hybrid learning.

equipment) that introduce the young to adult skills. Many games practice crucial components of adult skill (for details, see the companion paper).

In my view, then, hybrid learning in an appropriately organised learning environment is necessary for high fidelity high bandwidth information transmission. Indeed, as Avital and Jablonka illustrate, organised learning environments (though not as elaborately organised as those of the last 100, 000 years) is sufficient for the transmission of skill, though with a lower bandwidth and less fidelity. Avital and Jablonka point out that an ecological innovation (as in the Japanese macaque potato washing case) can re-organise the learning environment of the next generation. The change in parental lifeways changes the environment that the next generation explores, and changes the characteristic pattern of adult activities. The initial innovation may have been a fluke, but if it re-organises the way the innovators live, reacquiring that innovation in the next generation can become routine (Avital and Jablonka 2000; Jablonka and Lamb 2005). The moral is that amplified social learning and informational co-operation does not have to wait for special cognitive adaptations: we can get more social learning (via these hybrid routines involving organised learning environments) for free, if there are appropriate changes in ecology and social dynamics. Once hominins lived in social worlds in which cross-generation cultural learning played a crucial role in development, there was then selection for individual cognitive adaptations which improved the reliability and fidelity of learning.

So while Avital and Jablonka are right in thinking that the social transmission of innovation and skill need not require special cognitive adaptations in either the information source or soak, the reliability, bandwidth and fidelity of such transmission is enhanced by such adaptations. Very likely, individual adaptation for cultural learning are necessary for the transmission of skill sets as demanding as ours (Tomasello 1999; Alvard 2003). There has been important new work on the nature of these individual adaptations. Perhaps the best known and most influential is the line of work deriving from Michael Tomasello, arguing that the capacity for imitation learning plus mechanisms of shared attention and shared intentions (hence motivational as well as purely cognitive changes) play a fundamental role (Tomasello, Carpenter et al. 2005; Tomasello and Carpenter 2007). More recently, Gergely Csibra

and Gyorgy Gergely (with various coworkers) have developed a variant model of these adaptations that dovetails in two respects with the model of cultural learning defended here (Csibra and Gergely 2005; Gergely and Csibra 2005; Gergely and Csibra 2006). First, the Csibra-Gergely model is focused on the interaction between source and soak: both are adapted for the cultural transmission of information. Second, in the paradigm operation of the “pedagogy complex”, ecological and cultural information are learned simultaneously, and in a mutually supporting way: novel artefacts or acts are learned from demonstrations in combination with their linguistic labels and with the establishment of joint attention. Csibra and Gergely think that this model operates after source and soak exchange cues that establish that the source will demonstrate new and relevant information. So the capacity to learn from demonstration is important, but it is part of a complex which includes a sensitivity to cues of a teaching context (eye contact; motherese, imitation games and the like); the capacity to interpret directional gestures as reference fixing devices; the expectation that the directional marker will be followed by new and relevant information about that referent. Moreover, young children have the capacity to extract information rapidly in this context and detach it from its source: instead of learning that Alison likes broccoli⁹, the child learns that broccoli is good (Gergely, Egyed et al. 2007). In extracting information from these interactions, children act on the tacit assumption that information sources are benevolent, reliable and typical. This assumption is adaptive: for as we have seen, information of the kind that is acquired in these contexts — culturally specific information about objects, practices and their names — do not pose deception problems.

One of our adaptations for cultural learning is folk epistemology. Folk epistemology does play the anti-viral role Sperber identifies. But it also plays a role in folk education theory: people use their understanding of minds and representations to enhance the flow of expertise across the generations. In behaviourally modern human communities, successful life requires a raft of many skills, some of them complex. Their acquisition depends not just on socially assisted learning, but on high-density high-fidelity socially assisted learning. Apprenticeship learning is the model I have in mind. In preindustrial societies craft skills were acquired by long formal

⁹ Still less, instead of learning that Alison wants me to think she likes broccoli

apprenticeships, and many other skills were acquired by informal versions of the same situation: learning by doing, but doing that was supervised by experts, in part to increase expertise. The doing in part consisted in aiding the expert; in part consisted in working with supervision and advice in an environment rich in demonstrations and exemplars of good practice.

A behavioural program is the organization of a capacity into an interacting sequence of sub-capacities. In writing of the evolution of technical expertise, Richard Byrne has argued great apes learn “behavioural programs”, and to some extent, can learn them by observing others (Byrne 2003; Byrne 2004). Learning a behavioural program, even learning a behavioural program by imitation, is in no sense a metarepresentation skill. That is not true of teaching: effective teaching by demonstration requires an agent to make that program overt, and that requires models to represent their own capacity. More generally, adapting the learning environment of the inexpert requires theory of mind and other metarepresentation skills. The active supervision of learning requires the expert to understand what the inexpert can and cannot do, so they can assign tasks (and suggest exemplars) that lie within the inexpert’s capacities (but which stretch or consolidate them). In formal apprenticeship learning situations, they order skill acquisition, so that each step prepares the next; the same is likely to be true in many informal apprenticeships. The expert need not explicitly teach in order to improve the fidelity and reliability of skill acquisition. Task assignment; the provision of exemplars and examples; ordering trial and error learning problems so that each task prepares for the next improve the reliability and fidelity of learning, without requiring explicit teaching. Learning by doing can be adaptively organised by the expert, even without overt teaching. But in some cultures and contexts, the expert demonstrate their expertise, both in utilitarian activity and as demonstrations. Expert performance is often rapid and fluent, without obvious components. It is hard to learn from such performance unless the task is overtly decomposed into segments, each of which can be individually represented and practiced. Fluent natural performance is often less useful as a model than performances which are stylised (and accompanied by a meta-commentary). Such a stylised performance, of course, requires the model to represent to themselves their own competence.

Folk epistemology is important for the soak, too. They need to understand that a performance is stylised; that a crucial step has been slowed down, exaggerated or repeated, to make it more overt. As both Tomasello's joint attention and the Csibra-Gergely models make explicit, the soak and source need to read one another¹⁰. In short, agents exploit the resources of folk epistemology when and because they think about how to make what they know easier for others to learn, and when they adapt their own economic activities to compensate for their less skilled juvenile dependents.

If this is an important function of folk epistemology — if it is a learning and teaching tool, as well as anti-viral software — then the informational and technical complexity of expertise should connect to the capacity to represent one's own expertise. More complex technologies should map onto increases in an agent's capacity to represent the structure of their expertise; the high fidelity transmission of complex technique requires experts not just to organise their skill as a behavioural program, but to be aware of that program, and to be able to action elements of it independently. Likewise, complex technologies should map onto increases in an agent's capacity to choose helpful examples, and sequence learning tasks optimally: so they acquire and practice subskills in the right order. Likewise, complex technologies should map onto increases in an agent's capacity to identify just what is wrong with another's performance; not just to realise that something is wrong¹¹. This predicted connection between the increased complexity of expertise and self-reflective expertise is by no means trivial. As Herbert Dreyfus emphasised, much expertise is unreflective (Dreyfus 1992). So if this connection exists, it supports the idea that one function of folk epistemology is indeed to improve the efficiency of social learning.

This paper has focused on the mechanisms that maintain the complex web of information sharing on which behaviourally modern human life depends. The key aim

¹⁰ More prosaically, soaks need to choose their source, especially as they begin to explore their environment away from their immediate family. They need to be able to identify expertise. There is some evidence that quite young children have some sensitivity to varying levels of both epistemic confidence and performance ((Harris 2007; Jaswal and Malone 2007)).

¹¹ Of course, the expert are not on their own here. Just as expertise is a multi-generational invention, there will often be lore on the best practice for the transmission of expertise; lore which is itself a multi-generational construction, and which is itself an aspect of folk epistemology.

is to explain why that web does not decay in the face of failures to contribute, attempts to deceive and simple unreliability and to explain how it can be so powerful despite inevitable failures. There is a second project, equally important, of explaining the evolutionary construction of this complex. How did our lineage shift from the limited version of information-sharing that characterises great ape life to the complex, stabilised and obligate information-sharing of behavioural modernity?

In companion papers, I commit to a model of the evolution of ecological co-operation (which, of course, includes informational co-operation); a model that commits me to a coarse-grained perspective on this trajectory. For I have argued that co-operative defence against predation, co-operative foraging, especially male scavenging coalitions (but perhaps including a sexual division of labour) and reproductive co-operation are all relatively early forms of co-operation. More sophisticated technology, with a division of labour including specialisation, and an expansion of the resource basis of hominin life evolved more recently. Likewise, until relatively recently, there is no clear archaeological record showing that human groups existed as groups for themselves; as groups with a conscious self-identity. There is no problem of culturally learning the meaning of gang patches until there were gangs. On this view, I would expect co-operative signalling of information about the local ecological profile and short-term co-ordinating signals to be early forms of information sharing. The skill and technology base was still relatively simple, and hence the mechanisms Avital and Lamb identify might well be sufficient to explain their reliable cross-generational transmission. Information about skill and technology may well have leaked across the generations through cues and through parental impacts on the environment the next generation experienced rather than through explicit signals.

References

- Alvard, M. (2003). "The Adaptive Nature of Culture." *Evolutionary Anthropology* **12**: 136-149.
- Atran, S. (2001). "The Trouble With Memes: Inference versus Imitation in Cultural Creation." *Human Nature* **12**(4): 351-381.
- Avital, E. and E. Jablonka (2000). *Animal Traditions: Behavioural Inheritance in Evolution*. Cambridge, Cambridge University Press.
- Bacharach, M. and D. Gambetta (2001). Trust in Signs. *Trust in Society*. K. Cook. New York, Russel Sage Foundation. **II**: 148-184.

- Boehm, C. (1999). Hierarchy in the Forest. Cambridge, Mass, Harvard University Press.
- Boyd, R. and P. Richerson (1996). "Why Culture is Common but Cultural Evolution is Rare." Proceedings of the British Academy **88**: 77-93.
- Byrne, R. (2003). "Imitation as behaviour parsing." Philosophical Transactions of the Royal Society of London B **358**: 529-536.
- Byrne, R. (2004). The Manual Skills and Cognition that Lie Behind Hominid Tool Use. Evolutionary Origins of Great Ape Intelligence. A. Russon and D. R. Begun. Cambridge, Cambridge University Press: 31-44.
- Clark, A. (forthcoming). Supersizing The Mind. Oxford, Oxford University Press.
- Cosmides, L. and J. Tooby (2000). Consider The Sources: The Evolution of Adaptation for Decoupling and Metarepresentation. Metarepresentation. D. Sperber. New York, Oxford University Press: 53-116.
- Csibra, G. and G. Gergely (2005). Social Learning and Social Cognition: The Case for Pedagogy. Processes of Change in Brain and Cognitive Development. M. H. Johnson and Y. Munakata. Oxford, Oxford University Press.
- Danchin, E. and G. Luc-Alain (2004). "Public Information: From Nosy Neighbors to Cultural Evolution." Science **305**(July 23).
- Dennett, D. (1988). "Out of the Armchair and into the Field " Poetics Today **9**: 205-221.
- Dennett, D. C. (1983). " Intentional Systems in Cognitive Ethology: The "Panglossian Paradigm" Defended." Behavioral and Brain Sciences **6**: 343-390.
- Dreyfus, H. (1992). What Computers Still Can't Do: A Critique of Artificial Reason. Cambridge, MIT Press.
- Gambetta, D. (2005). Deceptive Mimicry in Humans. Perspectives on Imitation From Neuroscience to Social Science. S. Hurley and N. Chater. Cambridge, MIT Press. **2**: 221-241.
- Gergely, G. and G. Csibra (2005). "The Social Construction of the Cultural Mind." Interaction Studies **6**: 465-481.
- Gergely, G. and G. Csibra (2006). Sylvia's Recipe: The Role of Imitation and Pedagogy in the Transmission of Cultural Knowledge. Roots of Human Society: Culture, Cognition and Human Interaction. N. J. Enfield and S. C. Levenson. Oxford, Berg Publishers: 229-255.
- Gergely, G., K. Egyed, et al. (2007). "On Pedagogy." Developmental Science **10**(1): 139-146.
- Gil-White, F. (2005). Common Misunderstandings of Memes (and Genes): The promise and the limits of the genetic analogy to cultural transmission processes. Perspectives on Imitation: From Mirror Neurons to Memes. S. Hurley and N. Chater. Cambridge, MIT Press.
- Gurven, M., H. Kaplan, et al. (2006). "How Long Does It Take To Become a Proficient Hunter? Implications For The Evolution of Extended Development and Long Life Span." Journal of Human Evolution **51**: 454-470.
- Harris, P. L. (2007). "Trust." Developmental Science **10**(1): 135-138.
- Hauser, M. (1996). The Evolution of Communication. Cambridge, MIT Press.
- Henrich, J. and F. Gil-White (2001). "The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing The Benefits of Cultural Transmission." Evolution and Human Behavior **22**: 165-196.
- Hill, K. and H. Kaplan (1999). "Life History Traits in Humans: Theory and Empirical Studies." Annual Review of Anthropology **28**: 397-430.

- Jablonka, E. and M. Lamb (2005). Evolution in Four Dimensions. Cambridge, Mass, MIT Press.
- Jaswal, V. K. and L. S. Malone (2007). "Turning Believers into Sceptics: 3-Year-Olds' Sensitivity to Cues to Speaker Credibility." Journal of Cognition and Development **8**(3): 263-283.
- Kaplan, H., K. Hill, et al. (2000). "A Theory of Human Life History Evolution: Diet, Intelligence and Longevity." Evolutionary Anthropology **9**(4): 156-185.
- Laland, K. (2001). Imitation, Social Learning and Preparedness as Mechanisms of Bounded Rationality. Bounded Rationality: The Adaptive Toolbox. G. Gigerenzer and R. Selten. Cambridge, Mass, MIT Press: 233-248.
- Laland, K. and W. Hoppitt (2003). "Do Animals Have Culture?" Evolutionary Anthropology **12**: 150-159.
- List, C. (2004). "Democracy in Animal Groups: A Political Science Perspective." Trends in Ecology and Evolution **19**(4): 168-169.
- McBrearty, S. and A. Brooks (2000). "The Revolution That Wasn't: A New Interpretation of The Origin of Modern Human Behavior." Journal of Human Evolution **39**(5): 453-563.
- Richerson, P. J. and R. Boyd (2005). Not By Genes Alone: How Culture Transformed Human Evolution. Chicago, University of Chicago Press.
- Saunders, S. (forthcoming). "Costly Signalling: A Work in Progress." Biology and Philosophy.
- Sperber, D. (1997). "Intuitive and Reflective Beliefs." Mind and Language **12**: 67-83.
- Sperber, D. (2000). Metarepresentations in an Evolutionary Perspective. Metarepresentations: A Multidisciplinary Perspective. D. Sperber. Oxford, Oxford University Press: 117-137.
- Sperber, D. (2001). "An Evolutionary Perspective on Testimony and Argumentation." Philosophical Topics **29**: 401-413.
- Sterelny, K. (2003). Thought in a Hostile World. New York, Blackwell.
- Sterelny, K. (2006). "The Evolution and Evolvability of Culture." Mind and Language **21**(2): 137-165.
- Sterelny, K. (2007). "Social Intelligence, Human Intelligence and Niche Construction." Proceedings of the Royal Society, London (series B) **362**(1480): 719-730.
- Tomasello, M. (1999). The Cultural Origins of Human Cognition. Cambridge, Harvard University Press.
- Tomasello, M. (1999). "The Human Adaptation for Culture." Annual Review of Anthropology **28**: 509-529.
- Tomasello, M. and M. Carpenter (2007). "Shared Intentionality." Developmental Science **10**(1): 121-125.
- Tomasello, M., M. Carpenter, et al. (2005). "Understanding and sharing intentions: The origins of cultural cognition." Behavioral and Brain Sciences **28**: 675-691.
- Zahavi, A. and A. Zahavi (1997). The Handicap Principle: A Missing Piece of Darwin's Puzzle. Oxford, Oxford University Press.