The Fate of the Third Chimpanzee

The Nicod Lectures 2008

Kim Sterelny
Philosophy and Tempo and Mode
Victoria University of Wellington and the ANU

Session 3
Information Pooling and Its Evolution

I
Introduction

In the last two sessions, I have argued that (i) human decision making typically involves high load problems; (ii) that we often respond adaptively to such high load problems, even in novel environments; (iii) that we often are in such novel environments; (iv) and that our capacity to respond to novelty derives from our ability to preserve, accumulate and transmit cognitive capital. Human lifeways depend on the social mobilisation of the informational resources needed to meet the challenges of the different human lives. I have further argued that the transmission of these resources depends on some mix of domain specific biases (as in the relationship between norm learning and prosocial emotions); broad spectrum cognitive adaptations for social learning; and richly and specifically engineered developmental environments.

In this session, I focus on the evolutionary elaboration and stabilisation of information sharing at and across generations. To explain the evolution of this phenomenon, we need to identify an evolutionary trajectory, a sequence in which each link is a small variant on its predecessors. The sequence as a whole will take us from simple social learning of the kinds found in many great apes to our unique, elaborate and complex

forms of social learning. Moreover, co-operative cultural learning is both a special case of, and an important contributor to, our elaborately and inescapably co-operative lives. So since social learning is a form of co-operation, at each link in this sequence of increasing commitment to information sharing, we need to explain how information sharing generated an increasing profit, and why that profit is stable despite the threat of defection. As with other forms of co-operation, information-sharing seems to create an opportunity to gather the benefits of co-operation will avoiding its costs.

In the discussion to come, I make four central points.

(i) The evolution of high volume, high fidelity information sharing is not just the evolution of a particular kind of head. It is, as well the evolution of a particular kind of social milieu. High volume, high fidelity information flow depends on the evolution and stabilisation of social environments of a particular kind. Humans have not always lived in such environments, and hence human social worlds have not always been conducive to the accumulation of cognitive capital. So while individual cognitive adaptations are important, their evolution is by no means all we need to explain.

(ii) Dan Sperber and others have been right to emphasize that

information sharing is an instance of co-operation, and like other forms of co-operation, there are often possibilities of free riding and deception. He sees folk epistemology as evolved anti-deception technology. But while this threat is important, it is not uniform. There are (relatively) low-risk forms of informational co-operation, and the evolution of these can create an environment selecting for the further elaboration of information pooling. Less deception-vulnerable forms of informational co-operation evolve first, and their evolution scaffolds the evolution of anti-deception technology.

(iii) Because the threat of deception is not uniform, 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.

(iv) 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, read and assess signals sent through many channels; channels which vary in reliability, bandwidth, and in the kind of information that flows through them.

II
Behavioural Modernity

In the palaeobiological literature, there is a significant and growing literature on the so-called problem of behavioural modernity. This problem arises out of a mis-match between the biological origins of our species and the origins and establishment of characteristically human behavioural patterns. Anatomically modern humans appeared on the scene roughly 200k years ago. Yet these First Sapiens behaved (it seems) unlike any contemporary humans. Their material technology was much simpler; their foraging breadth was narrower; their social and cultural organization was more rudimentary. As far as I know, no-one doubts that there was a real contrast between First Sapiens and Moderns. But there has been considerable change in how that contrast is pictured. Until recently, it seemed as if the transition from First Sapiens to people like us was abrupt and co-ordinated. Somewhere in the band 50,000-40,000 kya, sapiens became human. Technology exploded, both in regional variation; in the size of individual toolkits; in the range of materials used, and in the complexity of individual tools. At the same time, the economic base of human life became broader. A wider range of animal species was taken, including those that are difficult or dangerous. Grains were gathered and ground to

flour. Marine resources were added to the human menu, and long distance trade networks were established. These economic and technological changes were coupled to changes in how humans conceived of themselves in their world. For decoration, ornamentation, and (a little later) musical instruments, cave art and figurines appear. This pulse seemed so dramatic that some suggested that it had to be the result of some final cognitive breakthrough in the biological organization of our minds.

More recently, researchers have suggested that many of the traits supposedly definitive of this revolution appeared earlier in Africa. These recent papers all suggest that the gap between earlier and later humans is not quite so dramatic as a simple reading of the record suggests. The same themes come through repeatedly. Features of material culture and foraging capacities that were once though to be diagnostic of Modernity turn out to have anticipations in the Middle Stone Age (i.e. roughly 285,000 to 50,000 years bp). So there are Middle Stone Age examples of stone blades, hafted tips, and even standardised tool shapes. There are very old spears from Germany, showing projectile technology, and bone tools of varying levels of sophistication. In short, there seems to have been many early anticipations of, and a slow build-up to, the establishment of behavioural modernity. The transformation is real, but not the result of dramatic change within a single lineage in a single

region. The puzzle has shifted, but it is still a puzzle, if we think that ancient sapiens had essentially the same cognitive horsepower as those of the last 50,000 years. Why did these humans take so long to generate the material and informational technology that is such an evident and dramatic feature of the last 50,000 years?

In answering this question, I think the Australian case is very instructive. The initial expansion of humans into Australia took place about 45,000 years bp. It could not have been accidental. There were too many water-crossings for anything remotely resembling the "pregnant women on a log" scenario to explain human arrival in ancient Australia. These humans were genetically modern, and they must have been in crucial respects cognitively modern too: they had technology complex enough to cross significant stretches of ocean. However, until 20k years or so ago, the Australian archaeological record resembles that of Middle Stone Age Africa. For the first 25,000 years, early Australians seem to have a limited technological toolkit; exploited a narrow resource band, and showed very limited signs of symbolic culture. The usual archaeological signatures of behavioural modernity emerged only in the last 20,000 years. We then see evidence of broad-range foraging; environmental management; technological innovation; obvious symbolic culture.

Jim O'Connell and Jim Allen take the Australian case to show that people can be behaviourally modern without showing that they are behaviourally modern. The first humans to reach Australia needed planning and innovative technology to arrive and establish. So O'Connell and Allen argue that for many thousands of years, Australians were behaviourally modern without seeming to be behaviourally modern. They do not consider the idea that Australians ceased to be modern after they arrived. Neglecting this possibility makes sense if we think modernity is coded and canalised in individual genomes. But it makes no sense if behavioural modernity is partially constituted by the organization of social life. Yet human cognitive skills depend very heavily on the epistemic technology and communal information resources to which we have access, and that access is sensitive to demography and ecology. That might have changed fundamentally as small numbers of people dispersed into an enormous landscape. There is no reason to assume that behavioural modernity is a fixed and genetically canalised feature of individual phenotypes; and that once behaviourally modern, always behaviourally modern. For behavioural modernity may well depend on an interaction between individual phenotypes and social environment.

Indeed, I shall defend the idea that behavioural modernity is a stabilised system of interaction between individual agents and their social environment. Specifically, it is the stabilised interaction pattern that makes the accumulation of cognitive capital not just possible but reliable. There is an important distinction between the conditions that allow information to be preserved reliably, and those that allow it to be expanded reliably. This difference allows us to make sense of the hominin record, which falls into three phases: a long phase of mere preservation; a not-yet-stable shift to expansion, and a final phase in which innovations and additions to the communal stock of information are much more reliably transmitted to the next generation. Thus hominin history began with a very long phase of technological conservatism. Long periods of technological, ecological and cultural stasis are punctuated by shifts to more complex technologies Simple chopping tools and flakes emerge approximately 2.6 million years ago in Africa and make a first appearance in Europe some time later. At about 1.6 million years ago, this technology is supplemented with the classic Acheulian handaxe. These are bi-facially flaked, and often have a standardised "tear drop" shape. Middle Stone Age points begin to appear about 280k years ago, and this change signals the arrival of hafted rather than hand-held tools.

From about 200,000 years ago, technological and ecological traditions become less conservative. There were innovations in this period which anticipated later technological revolutions, but often these innovations seem to fade out. The accumulation of innovation is not yet stable. The final phase, of course, is the signature period of behavioural modernity: innovation, regional variation, and expansion into all but the most forbidding habitats and inaccessible regions.

This overall pattern records the shift from one mode of cultural transmission to another: from transmission being reliable enough and of high enough fidelity to be able to preserve key informational resources of a community to transmission being sufficiently reliable and accurate to allow informational resources to be accumulated and transmitted. Bandwidth and fidelity improve. Accumulation demands both fidelity and bandwidth. Accumulation requires innovative small changes on established practices to be transmitted, not just the base practice itself. And it requires an increase in the volume of information that is transmitted. These require both individual cognitive adaptations and the right social environment. In the intermediate period (I suspect) the individual cognitive adaptations have evolved, but are probably not yet fine-tuned. Moreover the social and developmental environment necessary to accumulate cognitive resources has not yet stabilised.

III
Behavioural Modernity and Fidelity

Individual cognitive adaptations for cultural learning are important, but these evolve after the establishment of hominin lifeways dependent on cultural transmission. Avital and Jablonka show that some information can be created and preserved without specific adaptations. Traditions based on social learning can be stabilised by niche construction. An animal innovates successfully. As the result of that innovation, the animal's life ways are re-organised. The resource to which it now has access plays a central rather than a peripheral role in its ordinary ecological life. As a result, in those social species in which the offspring accompany their mother, ordinary exploration and trial and error learning will give the young many opportunities to learn to exploit the new resource. The initial innovation may have been a low probability event, but the transmission of the skill can be very probable, without any need to invoke high-cost cognitive adaptations.

The Avital and Jablonka model fits Oldowan technology quite naturally. A successful innovation by a single individual or small group sparked a local re-organization of their lifeway around the new resource. That change automatically re-

organised the learning environment of the next generation. So some accumulation and preservation is possible without specific adaptations for hi-fi social learning, so long as the social environment is friendly to the transmission of the new skill. The young have to stay with their parents, to be tolerated in close proximity while they play with what their parents use. Perhaps they also need to be inquisitive about what their parents are up to. But we do not need to suppose that adaptations for social learning preceded early, stable but simple and low bandwidth technological traditions. Rather, their establishment via niche construction created the selective environment favouring those adaptations.

However, once these lifeways establish and become typical for the species, this sets up new developmental and selective environments. The initial shift to a stone-technology based lifestyles depended on pre-existing mechanisms of adaptive plasticity; pre-existing potentials for manual dexterity; pre-existing foraging patterns. Once established, the new lifestyle will select for genetic variants that enable these new skills to be acquired with high reliability and low cost (it is easy to lose eyes and fingers flint-knapping). Avital and Jablonka's ideas show that in an appropriately organised learning environment, agents do not need individual adaptations for social learning to learn socially. Peter Richerson and Robert Boyd have developed

a related line of thought. They doubt that our adaptations for social learning are high fidelity mechanisms, and argue that the social environment compensates for low fidelity through redundancy. Naïve agents have many opportunities to acquire specific skills and critical information, and they develop models to show that redundancy — for example, a naive agent using many models rather than a single model — can compensate for low fidelity one-on-one learning. Thus so long as there is sufficient redundancy, a population can preserve its informational resources in transmission to the next generation through low fidelity channels.

So redundancy together with low fidelity transmission can preserve informational resources, allowing already established and widespread skills to be copied via multiple trials to the next generation. However, as Michael Tomasello argues, such mechanisms will not allow small, incremental improvements to existing techniques to be preserved, copied to the next generation, and spread to be the foundation for further improvement. For this reason, it is clear that the cultural learning characteristic of the transition to behavioural modernity and of later periods of human culture requires both individual cognitive adaptations for cultural learning and highly structured learning environments. For these social worlds depended on

both a large bandwidth and sufficient accuracy for a ratchet of improvement.

As I argued in session 1, apprentice learning offers a good general model of this combination. A skilled cabinet maker (for example) has absorbed an enormous amount of information and skill from his/her teachers. An apprentice obviously brings to the learning environment a complex set of individual cognitive adaptations: physical skills, theory of mind, joint attention, conditional reasoning, observation learning. Most apprentices acquiring complex skills benefit from explicit advice and instruction, and the observation of expertise in action. Often, those learning share information too, about both failure and success. But most learning is hybrid: apprentices mostly learn through socially structured trial and error learning. They learn on the job, but they are assigned jobs by those who understand how much or little they can do. So their trial and error learning often involves structured trials. Apprentice learning systems combine high fidelity with large bandwidth. These systems depend on population structure, not just individual cognitive adaptations. The size and organization of the local community is extremely important to its capacities to accumulate new information, and to preserve those resources. In particular, size helps. (i) Redundancy plays a critical role in buffering the group's informational resources. Larger groups store

information in more heads than smaller ones. Information can easily drift out of a small group, through unlucky accidents to those with rare skills. In addition though, as we saw in discussing Boyd and Richerson, redundancy plays an important role in compensating for low fidelity cultural learning. (ii) Second, in larger groups a larger market size allows more specialisation and more division of labour, both of which impact positively on a group's informational resources. (iii) Finally, all else equal, a more diverse group with a varied skill set is more likely to innovate than a small, more homogeneous group.

The transition to behavioural modernity was not, of course, the result of the formation, 50,000 years ago, of a Palaeolithic equivalent of medieval craft guilds. But I do think that the information rich, expertise dependent, forager lifestyles of this phase of human life depended on a similar combination of the organization of learning with specific adaptations for social learning. The persistence of these lifeways depended both on models sharing their expertise and on the reliable replication of the learning environment in which crucial expertise was acquired. Only thus can cognitive capital be accumulated; only thus did we become behaviourally modern.

Moreover, 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. For the flow of information does not depend chiefly on low-cost, arbitrary signals. Language is a superb tool of deception because linguistic symbols are arbitrary, and reference is displaced in space and time. In contrast, observation learning is a channel with 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 we shall see, these are honest signalling mechanisms. I noted earlier that Sperber sees solving the defection problem as central to understanding the evolution of co-operation. To that issue, we now turn, but with the important preliminary result that a critical form of cultural learning depends on informational channels which are relatively safe from deception.

IV
Sperber's Dilemma

We are habitual and obligate participants in rich networks of social or cultural learning; we are both information soaks and information sources. As with other forms of co-operation, sharing information is both risky and potentially profitable. Listening to another agent seems to offer the opportunity to acquire at negligible cost crucial information. 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 catastrophic. 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 soak 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 soak via deceptive signals. Dan Sperber has argued that our folk epistemology — our set of tools for representing and evaluating signs and signals — is a response to the problem of deception. Metarepresentation evolves as part of a mechanism of indirect scrutiny; of folk logic. We do not just represent representations; we assess them. We scrutinise 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. 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.

Thus Sperber proposes an antiviral model of folk epistemology. I think there is something importantly right about this idea. The problem of trust is genuine and ancient. Free-riding and deceptive manipulation is a risk to those engaging in information sharing. Moreover, 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 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. So I agree that Sperber's dilemma is important, but do not think it is ubiquitous.

V
My Nipples Explode With Delight

In my view, then, defection is a threat to some information sharing transactions but not others. The profit of cultural learning varies, as does its risks. They are contingent on the identities of source and soak; on the domain about which communication takes place; and on the communication channel. 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 future actions. The 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. Folk epistemology is a fallible tool, but it is the best agents have in managing and assessing conspiratorial whispers.

Diplomacy is indeed a stereotype of one form of cultural learning and communication, and in such cases folk epistemology does indeed play the role Sperber identifies. But consider a contrasting example from my youth: Monty Python's celebrated Hungarian Phrase Book sketch, in which a publisher produces a supposed 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". 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. But in such a case there could rarely be a

temptation to deceive. 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 information channel is the same as those of *Diplomacy* conspiracies. But with the change in topic, and with the change to a multi-agent, epistemically heterogeneous audience, the threat of deception vanishes. Not all phrase books are well-designed, but we discount the problem of manipulation for good reason.

In the appendix, I have represented the complexity of human cultural learning through two tables. These tables are not complete, but they highlight the complexity of human cultural learning, and the subtle interplays between cost, benefit and reliability. The next three sections illustrate these interactions.

VI
Sharing Ecological Information

Consider, first, the complex of issues around shared information about the local environment. 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, 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 valuable but apparently evolutionarily fragile. For the information channels are not intrinsically reliable. The information flow depends largely on language, gesture and similarly low-cost signals about spatiotemporally displaced targets of inquiry. Since the information flow concerns the elsewhere and the elsewhen, their veracity cannot be checked directly, against the world. And if temptations to deceive exist, the characteristics of the channel itself will not prevent agents succumbing to those temptations. If the profit of co-operation is generated by reciprocation over time, yet on particular occasions of information donation there is a temptation to defect, the problem of trust seems serious. The problem seems especially serious as the signals conveying information are not intrinsically reliable. How do agents police fair reciprocation of informational favours?

Despite the problem of low intrinsic reliability, I do not think Sperber's Dilemma is especially pressing in these 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. The audience 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. 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 veil of ignorance that severs the planning connection between false signal and Machiavellian consequences. 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, the benefits of ecological information pooling do not depend on ultra-sophisticated communication systems. 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 information pooling.

VII
Redundancy

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. I saw the norms and rituals in use as well as by report. It would have taken a persistent and disciplined conspiracy (far beyond their organization talents) to practice a deception upon me. We can reply on shared information about norms, customs, symbols in part because soaks normally acquire information of this kind redundantly and multiply.

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.

The Condorcet 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. 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. A crucial aspect of such cases is that the profit of co-operation does not derive from serial reciprocation; it is immediate.

VIII
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, sexual negotiation is often a high stakes activity. So there are high stakes co-ordination and partnership decisions in small scale societies, and even in these worlds agents 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. 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 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.

IX
Metarepresentation and Social Learning

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. 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. If I am right about the role of hybrid learning in human cultural transmission, decomposing a skill into a behavioural program is likely to be very important to us. 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).

Learning by doing can be adaptively organised by the expert, even without overt teaching. Task assignment; the provision of exemplars and examples; ordering trial and error learning problems so that each task prepares for the next all improve the

reliability and fidelity of learning, without requiring explicit teaching. But in some cultures and contexts, experts demonstrate their expertise. 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 that 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.

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 the expertise of a culture should connect to individual's capacities to represent their own expertise. More complex technologies should map onto increases in individual capacity to represent the structure of 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 one agent's capacity to identify just what is wrong with another's performance; not just to realise that something is wrong. If there is a connection between the increased complexity of expertise and self-reflective expertise, that is by no means trivial. As Herbert Dreyfus emphasised, much expertise is unreflective. 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, and on how that complex web established. I have argued that some forms of information sharing were available early, and scaffolded the evolution of other forms, because they were both profitable and relatively safe from exploitation. These early-evolving forms of informational co-operation change the selective environment, allowing the evolution of a suite of individual adaptations for social learning and of engineered environments. Folk epistemology is one such individual adaptation complex, but I think of it as a multi-purpose tool for organising information flow rather than a policing mechanism.

Appendix

Content Domain	Source benefit/loss	Soak benefit/loss	Collective ben
Long-life information re local environment. It is possible, but sometimes expensive to acquire this information by individual learning.	Apparent relative fitness sacrifice	High benefits; crucial information.	Collective manages l environment n efficiently; buffering gr info resources
short-life information re local environment. It is usually possible but sometimes expensive to acquire this information non-socially. Time constraints might make it impossible	Apparent relative fitness sacrifice; but information pooling can increase reliability	Variable. Sometimes crucial information	Threat/opportu management; information pooling incre reliability
Skills	Apparent	Very high benefits;	Division

Often not possible to acquire skills by individual learning	relative fitness sacrifice; but between adults sometimes gain by increase in reliability of information pooling	crucial life skills not otherwise to be had	labour; benefit specialisation; buffering gr info resour extending platform further innovat
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 defection; poli norms
Second-party social information (boasting) Sometimes possible but expensive to acquire by individual	Source may benefit via beneficial relation with soak	Choosing social partners; high risk-high reward decisions	neutral

learning			
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 labour; collec decision makir
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, customs/norms may not adaptive
Public representational media Typically this information must be acquired culturally	Source gains new tools for influencing soak's behaviour	Soak gains cognitive & communication resources	Division labour; impro co-ordination; buffering gr info resources

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