

# The Adaptive Problem of Absent Third-Party Punishment

Gordon P. D. Ingram

Jared R. Piazza

Jesse M. Bering

Institute of Cognition and Culture

Queen's University

Belfast

BT7 1NN

Northern Ireland

Email: [gingram02@qub.ac.uk](mailto:gingram02@qub.ac.uk)

[Fax: +44 9097 1332](tel:+4490971332)

[Phone: +44 9097 1340](tel:+4490971340)

## Abstract

*Language is a uniquely human behaviour, which has presented unique adaptive problems. Prominent among these is the transmission of information that may affect an individual's reputation. The possibility of punishment of those with a low reputation by absent third parties has created a selective pressure on human beings that is not shared by any other species. This has led to the evolution of unique cognitive structures that are capable of handling such a novel adaptive challenge. One of these, we argue, is the propositional theory of mind, which enables individuals to model, and potentially manipulate, their own reputation in the minds of other group members, by representing the beliefs that others have about the first party's intentions and actions. Support for our theoretical model is provided by an observational study on tattling in two preschools, and an experimental study of giving under threat of gossip in a dictator game.*

**Keywords:** Evolution of Language; Gossip; Indirect Reciprocity; Reputation; Theory of Mind

Language makes humans unique. Other animals employ complex systems of communication (Hauser, 1997), but their communicative signals are indexical in reference (Deacon, 1997). As with non-verbal communication in humans, the reference of animal signals inheres in the drawing of attention to the presence of a stimulus. The stimulus referred to may be internal, such as dominance/submission displays in dogs (Lorenz, 1966, Figure 3) or emotional signals in humans (Ekman, 1999). Or the stimulus may be external, such as predator alarm calls in vervet monkeys (Cheney & Seyfarth, 1990), the waggle dance in honeybees (Dyer, 2002), or pointing in humans (Kita, 2003). Alone among animals, humans employ an additional, symbolic mode of communication in the form of language: human communicative signals do not always refer to stimuli directly, but may refer to other combinations of symbols (Deacon, 1997). The use of recursive symbols means that, in addition to drawing attention to a present state of affairs in the world, humans can use language to describe past states of affairs, predict future states, and fabricate counterfactual states. This important design feature of language is called *displacement* (Crystal, 1997, pp. 400–401, following Hockett, 1960).

Language has doubtless helped humans to produce great cultural achievements: information about past innovations, for example in the production of stone tools, can be stored in the form of language and fed into the development of new innovations. Less obvious, perhaps, are the adaptive problems that language has created. Displacement means that individuals have to be concerned not only with the reactions of direct witnesses to their actions, but also with the potential reactions of many other individuals to whom the witnesses may communicate information about their actions. In this paper, we present the hypothesis that another uniquely human competence – our highly developed theory of mind – evolved in response to this adaptive problem.

We argue that language enabled the development of systems of indirect reciprocity, in which absent third parties punished individuals for negative behaviour towards others. We briefly present two recent studies by the authors that offer supporting evidence for this hypothesis. Young children's language when they report other children's behaviour shows design features,

such as honesty and negative bias, which ensure that absent third-party punishment is effective. And in adults, selfish behaviour in the dictator game is inhibited by the threat of gossip about such behaviour. We surmise that a sophisticated, propositional theory of mind is responsible for implementing this inhibition.

## **Gossip and indirect reciprocity**

Recently, two important areas of scientific research have opened up in relation to the evolution of uniquely human behaviour. One area of research has focused on the role of language in the expansion of human group sizes; the other has examined the theoretical conditions necessary for the evolution of cooperation in large groups. Our main aim in this section is to show that these two independent areas of research are systematically related. As Nowak and Sigmund (2005, p. 1295) put it: “The co-evolution of human language and cooperation by indirect reciprocity is a fascinating and as yet unexplored topic.”

Dunbar (1993; 2004a; 2004b) proposes that a propensity towards a specific kind of language – namely, gossip about other social agents – enabled the evolution of complex human societies. The size of other primate societies, according to Dunbar, is limited by the amount of time that they can devote to the practice of social grooming, which serves to reinforce social bonds and communicate emotional information between allies. In humans, a selective pressure towards larger group sizes led to the evolution of language as an adaptation which could perform similar functions to grooming, in maintaining group cohesion, but which was better suited to being deployed in a distributed manner across a large social group. This was partly because language enabled our ancestors to communicate social information to more than one individual at a time, and partly because it enabled our ancestors to communicate strategic information about the behaviour of absent third parties. In this chapter we focus on the latter activity, which we consider to be broadly synonymous with gossip.

Communication about absent third-parties allowed our ancestors to stay in touch with group members who were physically distant, enabling early human groups to spread out over far wider areas than were occupied by other primates, and enabling individuals to maintain wider networks

of contacts. Such communication would also have helped to encourage cooperation within groups by spreading information about free riders and other norm violators (see also Dunbar, 1999). In Dunbar's words:

[Language] allows us to exchange information about other people, so short-circuiting the laborious process of finding out how they behave. For monkeys and apes, all this has to be done by direct observation. I may never know that you are unreliable until I see you in action with an ally, and that opportunity is likely to occur only rarely. But a mutual acquaintance may be able to report on his or her experiences of you, and so warn me against you – especially if they share a common interest with me. Friends and relations will not want to see their allies being exploited by other individuals, since a cost borne by an ally is ultimately a cost borne by them too.

(Dunbar, 2004b, p. 79)

A similar argument is made by Boehm (1999; 2000) in relation to the evolution of egalitarianism in humans. Consider the case of a strong individual living in a group of social primates that bullies a weaker group member, for example by snatching food from its grasp. In most species the weaker individual has only two possible avenues of resistance:

- i.* It may withhold future cooperation from the aggressor
- ii.* It may attempt to alert other individuals *in the vicinity* with a vested interest in its fitness, which may be powerful enough to deter or punish the aggressor.

Language dramatically expanded the scope of the second form of resistance. Early humans did not have to rely on a powerful third party directly witnessing an act of aggression: the victim themselves, or any other witness, could spread information about the aggression around an entire social network. Furthermore, third parties who heard about the aggression did not have to punish the aggressor single-handedly, but could use language to recruit others to support them, as well as to justify to others the action that they took. Boehm (1999; 2000) argues that the use of language

to punish or deter aggression led to the evolution of human (hunter-gatherer) societies that were highly egalitarian compared to the rigid hierarchies of chimpanzee societies.

Another way of looking at the same process is in terms of *indirect reciprocity* (Alexander, 1987; Nowak & Sigmund, 2005). Systems of direct reciprocity (Trivers, 1971) rely on the return of a positive or negative payback for an action to the individual *A* who performed that action, by the recipient *B* of the action. The potential of direct reciprocity for enhancing cooperation is limited by the scope for future interactions between these two individuals *A* and *B*. In contrast, in systems of indirect reciprocity payback may be returned either by the recipient *B*, by those with whom *B* shares genetic material, or even by unrelated group members who learn about the incident (Alexander, 1987, p.85). Indirect reciprocity can therefore lead to cooperation even in cases where *A* and *B* are unlikely to meet again. While direct reciprocity may be exhibited by some (though perhaps not many) other animals, indirect reciprocity appears to be exhibited only by humans (Alexander, 1987; Richerson, Boyd, & Henrich, 2003, p. 379). It seems plausible that systems of indirect reciprocity are responsible for the unique ability of humans to cooperate in large groups of genetically unrelated individuals (but see Richerson, et al., 2003, for a different view). We contend that the novel possibilities offered by indirect observation through language greatly expanded the scope of indirect reciprocity for maintaining cooperation.

Surprisingly, theorists have often neglected the importance of language in potentiating systems of indirect reciprocity. For example, Alexander (1987, p. 77) writes that: "Indirect reciprocity develops because interactions are repeated, or flow among a society's members, and because information about subsequent interactions can be gleaned from observing the reciprocal interactions of others." Yet plenty of non-human animals are capable of distinguishing between individual group members and engaging in repeated interactions; so this does not explain why other species have not evolved cognitive mechanisms for deciding whether conspecifics would make suitable cooperative partners. The key word in the preceding quotation is "observing," referring to the observation of an interaction between *A* and *B* by a third party, *C*. Alexander fails to note that in most cases, only a small and not necessarily representative proportion of *A* and *B*'s

interactions will be directly observed by *C* (Nowak & Sigmund, 2005). In humans, however, *C*'s observation is often indirect, based on the testimony of *A* or *B* (or indeed a direct observer, *D*). Linguistic testimony greatly increases the observability of interactions, and hence the scope of indirect reciprocity.

Indirect reciprocity is generally mediated by *A*'s *reputation* within a social network – where reputation may be defined as the information that is known about *A*'s behaviour and personality at second hand, independently of direct experience. Over the last decade there has been increasing interest in formal models and experimental paradigms in which reputation encourages the evolution of cooperation (e.g., Nowak & Sigmund, 1998; Semmann, Krambeck, & Milinski, 2004; Wedekind & Braithwaite, 2002; see Nowak & Sigmund, 2005, for a review). These models incorporate the important point that individuals in a society of cooperators are generally in competition with each other through their reputations (see especially Barclay & Willer, 2007). For instance, in Nowak and Sigmund's (1998) model an individual agent's "image score" – which is analogous to a simple reputation – controls that agent's access to mutually beneficial interactions. Researchers in this area sometimes note in passing that reputations are constructed and disseminated via language (but see Enquist & Leimar, 1993, and Nowak & Sigmund, 2005, for fuller treatments). Yet for modelling purposes, such research often treats image scores as though they were stamped on individuals' foreheads, ready for potential exchange partners to read instantly and unambiguously. Of course, in the real world, reputations are often manipulated via exaggeration, innuendo or downright deception, and getting to the truth of the actions and personalities behind them can be very difficult. Humans constantly use language to tell people about our own and others' actions; and we frequently use language to make value judgements about our own and others' actions (Foster, 2004).

Another line of research (e.g., Boyd, Gintis, Bowles, & Richerson, 2003; Fehr & Gächter, 2002; Fehr & Fischbacher, 2004; Fowler, 2005) has focused on altruistic punishment as a key mechanism in the evolution of human cooperation, and as an alternative to reputation-based mechanisms. However, the dichotomy between punishment and reputation may be illusory,

because in human societies those who punish a transgression are not always (or even often) direct witnesses to that transgression. Most punishment of transgressors – including punishment administered by the legal systems of complex societies – relies on the spread of information by witnesses via language. In other words, it relies on second-hand information that lowers a suspect's reputation. It is important to remember that indirect reciprocity can be negative as well as positive, which means that “altruistic punishment” may be considered to be a form of negative indirect reciprocation, and one which may be expected to bring its own rewards through positive indirect reciprocity (i.e., it is also a form of costly signalling; see Smith & Bliege Bird, 2005).

As with models of indirect reciprocity, the importance of language has been neglected or even denied in models of altruistic punishment. Richerson, Boyd and Henrich (2003) argue that indirect reciprocity, as mediated by language, could not have been solely responsible for the evolution of large-scale cooperation in humans, because it is not easy to distinguish true gossip from false:

It is not obvious that language potentiates indirect reciprocity.

Whereas superficially language may seem to promote the exchange of high-quality information required for indirect reciprocity to favour cooperation, this addition merely changes the question slightly to one of why individuals would cooperate in information sharing; language merely recreates the same public goods dilemma. Lies about hunting success, for example, are difficult to check and often ambiguous. Among the Gunwinggu (Australian foragers), members of one band often lied to members of other bands about their success to avoid having to share meat (Altman & Peterson, 1988).

(Richerson et al., 2003, p. 257, n. 3)

Richerson and his colleagues refer to this line of reasoning as the “cheap talk” argument. However, their argument has a number of weaknesses. Firstly, verbal deception is only cheap if it

goes undetected; if detected, it may incur penalties that are just as severe as those for stinginess or other forms of cheating. For example, in a recent experimental study, Tyler, Feldman and Reichert (2006) found that frequency of lying by confederates is inversely related to judgements of likeability that participants make about the confederates. Therefore lying may foreclose opportunities for cooperation by advertising the liar's unreliability as a cooperative partner. Secondly, while some lies may be hard to detect, others are not. If three or more people witnessed an event, and one person's story differs from everyone else's, that person risks being labelled a liar. Since it is not always easy to know exactly how many people witnessed something happen, or who they may tell and get to repeat the story, honesty is usually the safest policy. Thirdly, as the example of the Gunwinggu shows, deception may be adaptive between groups but not within groups. In general, the likelihood of deception increases where there are conflicts of interest, such as between mimics and their predators (Alexander 1987, pp.73 – 75; Dawkins & Krebs, 1978). In an extremely social species such as humans, conflicts of interest are likely to be far stronger between groups than within groups. And while one hunting band may know little about another band's exploits, it would be much more difficult for an individual hunter to conceal success or failure from other members of the same band. Some degree of honest communication is surely necessary to coordinate the activities of a group such as a hunting band in the first place.

Despite its flaws, Richerson et al.'s (2003) argument does help to focus attention on one of the main characteristics that gossip needs to possess in order to support systems of indirect reciprocity. If gossip is to promote cooperation, it must be reliable: either it should be mostly trustworthy, or it should be easy to distinguish between trustworthy and untrustworthy gossip. The second possibility – that humans have reliable cognitive mechanisms for distinguishing between truthful and deceptive gossip – was investigated by Hess and Hagen (2006). In a series of experiments using social psychological methods, they found that participants were more likely to believe gossip if it came from multiple sources, and less likely to believe negative gossip if the source was someone who was in competition with the subject of the gossip. Their results suggest that people may be equipped with strategies for marking gossip as either reliable or

untrustworthy, just as they are equipped with mechanisms for spotting cheaters in the first place (Cosmides & Tooby, 1992). Taken together, these two sets of cognitive mechanisms would seem to encourage the growth of indirect reciprocity, by ensuring the extraction and reliable transmission of information about antisocial activities.

The evolution of language and the evolution of cooperation through indirect reciprocity are intimately linked, as theorists have recently begun to acknowledge (Nowak & Sigmund, 2005). Research into altruistic punishment reminds us that indirect reciprocity can be negative as well as positive. In fact, it seems that in certain contexts, the use of language to support indirect reciprocity may be biased towards achieving punishment rather than rewards. The next section presents evidence from one such context, namely social cognitive development in children.

### **Tattling: the reporting of negative behaviour by young children**

The development of gossip in young children has been little studied, but it is generally accepted that the behaviour of absent others is not nearly as common a conversational topic for children under the age of 5 as it is for adults (Engel & Li, 2004; Fine, 1977). Yet children do frequently report to adults the negative behaviour of other children who are close at hand – a kind of informal linguistic testimony known as *tattling*. Tattling is the act of informing on a peer's activities, with the intention that this will lead to his/her punishment. This linguistic behaviour can be seen as an attempt to secure absent third-party punishment, even though the audience is not completely absent, because the adult has often not witnessed the behaviour described, or at least was not paying attention to it.

Tattling is a frequent (and frequently annoying, for teachers) occurrence in the classroom – much more common than the reporting of positive behaviour by peers, which has to be actively encouraged by teachers if it is to take place with any regularity (Skinner, Cashwell, & Skinner, 2000). Young children's general bias towards reporting negative behaviour is consistent with Harris and Núñez's (1996; see also Núñez & Harris, 1998) finding that three-to-four-year-olds' deontic reasoning about actions that would breach a permission rule is superior to their indicative reasoning about actions that would breach a description rule. Harris and Nunez (1996, Experiment

4) found that their participants were better at identifying a picture where a child was “doing something naughty” than a picture where a child was “doing something different”. As they point out, this task is closely related to Cosmides and Tooby’s (1992) cheater detection task. Cosmides and Tooby famously demonstrated that performance on the Wason selection task is greatly enhanced by framing the task in terms of a possible breach of social contract (e.g. checking a drinker’s ID before serving them alcohol). When performing this sort of reasoning about social norms, fast and frugal cognitive mechanisms are activated that are not usually deployed when reasoning about abstract symbols. It seems that heightened reasoning about norm violation is reflected in young children’s everyday behaviour, as they are more likely to report breaches of the teachers’ rules than unusual actions by other children that are not in breach of a rule.

One of the few studies to focus on tattling itself is by Ross and den Bak-Lammers, who carried out a longitudinal observational study of tattling in the home environment between Canadian sibling dyads aged 2 and 4 (den Bak & Ross, 1996) and 4 and 6 (Ross & den Bak-Lammers, 1998). The authors found that: (1) reports of a sibling’s negative (that is, rule-violating) behaviour are much more frequent than reports of positive or neutral behaviour combined; and (2) tattling is overwhelmingly truthful. They also found that the proportion of negative reports decreased significantly as children grew older, suggesting important differences between tattling in children and the more inhibited forms of gossip practised by adults.

One of us (GI) recently conducted an observational study of tattling in the very different social setting of the pre-school. GI carried out research in two pre-schools in Belfast, with one group of 16 children and another group of 24 children. Both groups were aged 3–4 years, with mean ages of 4 years 1 month and 4 years 2 months respectively. A mixture of participant observation and behavioural ecological techniques (event sampling, social network analysis, and focal follows) were used to analyze the content and context of children’s reports of peers’ behaviour. These two types of methodologies complemented each other well: participant observation provided qualitative insights into the meaning and motivation of tattling, while the

quantitative techniques allowed for a statistical analysis of the defining characteristics of tattling events.

GI spent a total of 66 hours engaged in event sampling in the two schools, recording 391 events in which children reported the behaviour of peers (Ingram, 2007; Ingram & Bering, forthcoming). Mean percentage results for the valence, and truth value of tattling in both pre-schools are summarised in Table 1, with sibling data from Ross and den Bak-Lammers (1998) for comparison:

< --- *INSERT TABLE 1 HERE* --- >

In all six contexts, the clear majority of young children's behavioural reports concern some sort of proscribed behaviour. The most common types of transgressions reported are property disputes and instances of physical harm – although in classroom settings, where there are many more or less arbitrary rules, breaches of social convention were frequently reported too. The finding that the content of peer reports is more likely to be negative in the pre-school setting than in the home setting could be due (at least in part) to sampling bias, because complaints about negative behaviour would be easier to notice than comments about positive or neutral behaviour in the noisy environment of the pre-school.

The truthfulness of children's behavioural reports is even more striking. It is noteworthy that most of the tiny minority of false reports recorded by GI seemed to be accidentally rather than deliberately false. It was impossible to know for sure whether a child was lying, but false tattling seemed to be more often the result of one of three different types of mistake: a misidentification of the child who was reported on; a bias towards attributing hostile intentions (e.g. "He pushed me" when another child accidentally bumped into the tattler – see Orobrio de Castro, Veerman, et al., 2002); or a misuse of language (e.g. saying "He punched me" but meaning "He pushed me"). There were less than instances of deception in all 391 recorded verbal reports.

The experimental literature on deception by children is equivocal (Reddy, 2007), with some studies suggesting that children as young as 2 or 3 are capable of intentional deception, whereas others indicate that it is not consistently achieved until the age of 4 or 5 (both sets of studies reviewed in Carlson, Moses, & Hix, 1998). In an attempt to resolve this contradiction, Carlson, Moses and Hix (1998) compared 3-year-old children's abilities to deceive by pointing with deception using a novel pictorial cue. They found that the children had much more difficulty with deceptive pointing, and argue that this is because children were unable to inhibit the accurate use of pointing as an action that has been reinforced through years of interaction with caregivers (and may even be instinctive).

A similar argument might be made for indicative uses of language to describe social events. Young children may not be able to inhibit their impulse to provide an accurate description of events, much as they might like to get a rival into trouble. It is interesting that the children in our study frequently *did* attempt deception in other contexts, for example by making vigorous (but plainly false) denials when they themselves were accused of a transgression (see also Lewis, Stanger, & Sullivan, 1989, who observed high rates of deception when 3-year-old children were asked whether they had peeked in a box that they were not supposed to look in). Perhaps the urge for self-preservation is enough to inhibit the impulse to tell the truth in such cases. Or it could be that preschool children do not tend to fabricate accounts of others' transgressions because their primary motivation in making such reports is not to get others in trouble without good reason, but to draw an authority figure's attention to authentic rule violations. A third possibility is that the executive demands of deception by simple denial of an accusation are much weaker than those involved in fabricating a story from scratch. Controlled experimentation would be required to test between these competing explanations.

Regardless of the proximate causes for the lack of deceptive tattling by children, it is tempting to postulate an ultimate cause. In small-scale societies, young children mainly communicate with members of their kin group, and this presumably reflects conditions in the ancestral evolutionary environment. The early preponderance of communication within the kin

group means that there are few conflicts of interest between tattlers and their audience. As Alexander (1987, pp.73–75) notes, a high frequency of deception is expected in any communication where there are significant conflicts of interest between the communicative partners (such as mimics deceiving predators; Dawkins & Krebs, 1978). Congruence of interests is easy to understand in the case of siblings tattling to parents, who are naturally interested in making sure that neither sibling is unfairly exploited by the other, because they have a huge amount of resources invested in both. Since young children are not prepared by evolution for life in the institutional environment of the preschool, it is reasonable to suppose that they treat staff at the preschool as surrogate parents, or at least as senior members of their kin group. Moreover, children use language not primarily to communicate useful information to arbitrary others within the social group, but far more often to communicate their own needs and desires to adults who look after them – and this is reflected in the egocentric nature of tattling, in which around 75% of reports were made about a transgression of which the tattler was the victim. It would make little sense for tattlers to lie in these circumstances: doing so could lead to a “cry wolf” type of scenario, in which the tattler is ignored just when they really need help. For language to be adaptive for a young child, it has to be an honest signal of need. This helps to ensure that reports of peers’ behaviour are reliable and are not undermined by the problems with “cheap talk” discussed by Richerson et al. (2003).

In summary, young children frequently report the behaviour of other children, and show a bias towards reporting negative, rule-violating behaviour. Tattling is adaptive at the individual level because it allows the absent-third party punishment of aggressors and transgressors who would not be susceptible to direct punishment by weak individuals such as children. Two important properties of children’s tattling – truthfulness and negative bias – are robust across varying social contexts, and would favour the development of systems of indirect reciprocity incorporating punishment of offenders by a group (or by those with authority within a group).

## **Experimental evidence of reputation management under threat of gossip**

The children in the observational study were too young to show much behavioural inhibition in response to the likelihood of tattling (although they did occasionally desist from an activity if directly threatened with tattling). In contrast, we hypothesized that an adult individual's selfish impulses are often inhibited by the risk that negative social information about his or her actions will be transmitted to absent third parties. As a preliminary test of this hypothesis, we employed an economic experiment called the *dictator game* (Camerer, 2003, ch. 2; Kahneman, Knetsch, & Thaler, 1986), in which each participant is asked to divide a monetary endowment between themselves and one other participant. Previous findings with the dictator game suggest that even in anonymous settings, equitable allocations are often the mode (Camerer, 2003, Table 2.4). However, selfish allocations tend to become more frequent as the "social distance" (or social isolation) between players increases, in highly anonymized experimental paradigms (Hoffman, McCabe, & Smith, 1996). Conversely, Bohnet and Frey (1999) found that dictators who could visually identify their recipients were less selfish with their allocations than dictators who could not identify their recipients, and selfishness decreased further when recipients were required to disclose their names. Even showing generic eyespots on a computer while dictators are making their decisions (Haley & Fessler, 2005) – or in a naturalistic setting, displaying an image of a pair of eyes rather than an image of flowers above a canteen "honesty box" (Bateson, Nettle, & Roberts, 2006) – can significantly increase allocations. The possibility of identifying the donor also has a strong effect on generosity. Burnham (2003) found that dictators' allocations were at similarly high levels – significantly higher than in an anonymized control – (i) when they were shown a photo of the recipient, and (ii) when they were told that their own photo would be shown to the recipient.

Given the importance of anonymity and identification in this research, we decided to investigate the role of language – that is, gossip – in facilitating the identification of individual actors and the spread of information about their actions. One of us (JP) used a modified dictator game to investigate the effect of two variables on altruistic behaviour (see Piazza & Bering,

forthcoming). The first variable (*reputational concern*) indicated whether or not a third party could identify the participant. This variable was manipulated using a confederate. Before being instructed about the dictator game, participants either did or did not meet and disclose personally identifying information (e.g., their name, place of study, place of residence) to a female confederate in private. The second variable (*threat of gossip*) indicated whether or not participants were told that the person directly affected by their economic decision (the receiver) would be discussing their economic decision with a third party. For some participants, this third party was the confederate with whom they had disclosed personally identifying information. For others, this third party could not identify them. The experiment thus had a 2 x 2 (*reputational concern* x *threat of gossip*) factorial, between-subjects design with *allocation scores* (i.e., number of tokens allocated to the receiver) as the main dependent measure.

The game itself entailed having participants unilaterally distribute an endowment of ten tokens between themselves and an anonymous second party (the Receiver). Each individual token represented a chance to have one's name entered into a lottery drawing to win £100. Participants were instructed that they would never meet the Receiver and that they were free to distribute the endowment however they wished – the only constraint was that the distribution total had to equal 10. Measures were also taken to reduce the motivation for building a reputation with the experimenter, in that participants' economic decisions were written down and enclosed in opaque envelopes to conceal them from the experimenter.

< --- INSERT TABLE 2 HERE --- >

The results were that participants who received the threat of gossip variable were significantly more generous with their allocations to the Receiver (see Table 2, and Piazza & Bering, forthcoming). However, simple-effects analysis revealed that this was only the case for participants who had previously disclosed personally identifying information to the recipient of

the gossip (the third party). When the recipient of the gossip was unable to identify the participant, participants in this condition were no more generous with their allocations than control participants (i.e., participants without the threat of gossip). This suggests that the threat of gossip can promote generous economic decisions, but only when an individual's reputation is placed at risk by the threat.

These findings are consistent with the hypothesis that selective pressure to present the self strategically (i.e., to promote a rewarding reputation) increased during human evolution with the emergence of language. This was accomplished by inhibiting selfish behaviour in order to prevent negative social information from being disseminated to absent third parties (i.e., potential mates and coalition partners). In the experiment, the threat of gossip appeared to inhibit selfish behaviour and encourage prosocial behaviour, but only when the threat was backed by real concerns about one's reputation. When the recipient of the gossip could identify the participant, the gossip prime led to a significant increase in allocations.

### **Perspective-taking and the management of strategic social information**

While the two empirical studies described in this paper used very different methodologies, their results are highly complementary. Our observational study shows that young children are biased towards the communication of truthful negative information about peers' activities, and that they use this to compete for adult attention. Tattling frequently results in punishment of those whose antisocial activities are reported. However, at the age of 3–4 years, the risk of absent third-party punishment does not seem to inhibit antisocial behaviour in general. This finding contrasts with Piazza and Bering's (forthcoming) results for the dictator game. In adults, the mere possibility of someone talking to a third party about a protagonist's activities, combined with the ability of the third party to identify the protagonist, is enough to increase prosocial behaviour in the dictator game. (Note that the sort of behaviour that the threat of gossip inhibits in the dictator game has a natural counterpart in the preschool environment, where a refusal to share is a frequent subject of tattling.) The question naturally arises: what are the developmental mechanisms by which the

threat of absent third-party punishment encourages the inhibition of antisocial behaviour, as children grow older?

Given the conflict between the risks of impairing one's reputation by acting selfishly and the costs associated with behavioural inhibition, early humans must have been under intense selective pressure to find a way of balancing these two forces. The solution, we argue, was to develop improved theory of mind (ToM) skills. Great apes and some other animals, such as social carnivores, may have a rudimentary ToM in the sense that they can predict behaviour by anticipating the intentions of other social agents (Premack & Woodruff, 1978; Tomasello, Call, & Hare, 2003). ToM in humans is vastly more sophisticated, however. It is possible that the divergence of improved human ToM skills is seen at a very simple level, in terms of an awareness that seeing leads to knowing (O'Neill, 1996), since evidence that chimpanzees possess this awareness is mixed (compare Hare, Call, Agnetta, & Tomasello, 2000, with Povinelli & Eddy, 1996; see also Povinelli & Bering, 2002). As discussed in the previous section, humans tend to behave more altruistically when observers are present (Bateson, Nettle, & Roberts, 2006; Burnham, 2003; Haley & Fessler, 2005). Linguistic testimony increases the number of potential observers of an individual's behaviour, thereby increasing the risk of punishment. There is no reason for a chimpanzee to inhibit its behaviour in response to the observation of a subordinate individual, and therefore little selective pressure for chimpanzees to respond to general cues of observation (rather than, for example, the presence of a dominant individual). Humans, on the other hand, must be sensitive to the presence of *any* possible observers of their behaviour, since it is quite possible for these observers to tell *anyone* else. Awareness of the link between seeing and knowing may have been the first stage in the development of improved perspective-taking skills in humans.

Whether or not humans show increased sensitivity to the link between seeing and knowing, it seems that a fully propositional theory of mind – which may logically require language – is unique to humans (Heyes, 1998). Propositional ToM is also known as the belief/desire model (Wellman, 1990), and includes a capacity to model false beliefs (Flavell, 1999; Wellman, Cross,

& Watson, 2001; Wimmer & Perner, 1983). The development of reasoning about beliefs and desires created the potential for us to model our reputations in the eyes of others, because in order to assess whether other people will see us in a good or bad light, we have to represent both what others believe to be the case about our actions, and what they desire our actions to be. In fact, because a representation of the actor's intentions is important for evaluating action, a fully developed model of reputation may require second-order theory of mind (Liddle & Nettle, 2006; Perner & Wimmer, 1985; Perner, 1988) – that is, an ability to think about what others believe our own intentions to be.

Recent research has highlighted the links between theory of mind, inhibition, and executive control (e.g. Carlson & Moses, 2001; Perner & Lang, 1999). All of these cognitive functions are associated with the prefrontal cortex (Gallagher & Frith, 2003; Stone, Baron-Cohen, & Knight, 1998), which is one of the brain areas that has shown the greatest expansion during human evolution (Deacon, 1997; Povinelli & Preuss, 1995; Schoenemann, 2006). It seems likely that propositional ToM is a vital part of the conscious mind, where it acts as a sort of “cognitive override” exerting inhibitory control over an individual's actions (Alexander, 1987, pp. 107–110; Bering & Shackelford, 2004). When deciding whether to behave selfishly or unselfishly, humans use conscious ToM skills to assess the likelihood of discovery, and also to estimate other people's approval or disapproval of their behaviour. They are thereby modelling the likely effects of their actions on their reputations – a process that may occur either consciously or unconsciously.

It has previously been suggested that ToM played a key role in the evolution of human cooperation (e.g., Bjorklund, Cormier, & Rosenberg, 2005; Dunbar, 2003; Povinelli & Preuss, 1995). The novel part of our argument is the drawing of an explicit connection between the evolution of language and the evolution of ToM. Language was not just a cognitive pre-requisite for propositional ToM to appear (as might be inferred from the fact that language is developed to a high degree in children well before they pass the false belief task<sup>1</sup>). Language was an *adaptive*

---

<sup>1</sup> See Astington, 2006, for a recent account of the complex interrelationship between language and theory of mind in child development.

*problem* for the individual because it frequently led to absent third-party punishment. This adaptive problem was solved by new adaptations, in the form of highly developed ToM and perspective-taking skills.

The position set out in this chapter is superficially at odds with the views of Nowak and Sigmund, who write that:

indirect reciprocity seems to require a ‘theory of mind’. Whereas altruism directed towards kin works because similar genomes reside in different organisms, reciprocal altruism recognizes that similar minds emerge from different brains. It is easy to conceive that an organism experiences as ‘good’ or ‘bad’ anything that affects the organism’s own reproductive fitness in a positive or negative sense. The step from there to judging, as ‘good’ or ‘bad’, actions between third parties, is not obvious. The same terms ‘good’ and ‘bad’ that are applied to pleasure and pain are also used for moral judgements: this linguistic quirk reveals an astonishing degree of empathy, and reflects highly developed faculties for cognition and abstraction.

(Nowak & Sigmund, 2005, p. 1291)

However, as discussed above, theory of mind is not limited to a propositional modelling of false beliefs, but should be considered as a complex portfolio of skills. Empathizing is one such skill, and probably arises earlier in development than does the representation of false beliefs (Flavell, 1999, citing Baldwin & Moses, 1994; Hoffman, 2000; Nichols, 2004, ch. 2). Some form of empathy may even be present in non-human primates (Preston & de Waal, 2002), dolphins (Kuczaj, Tranel, Trone, & Hill, 2001), and elephants (Douglas-Hamilton, Bhalla, Wittemyer, & Vollrath, 2006). In any case, early forms of indirect reciprocity may have been supported by the tattling of immature individuals to their parents, and it is doubtful that a fully empathic response is needed in order for a mother to defend her young. It is not difficult to imagine that early humans, possessing sophisticated enough brains to be capable of some sort of proto-language,

would also be capable of enough empathy (or something like it) to judge actions between third parties – especially where kin or close associates are involved – as ‘good’ or ‘bad.’

Finally, an important prediction of our evolutionary hypothesis is that ToM eventually feeds back into the linguistic behaviour by which strategic social information is spread. This is because of recursive, higher-order ToM skills: the ability to say to oneself, for instance, "she thinks that I think that..." (Perner & Wimmer, 1985). Using these skills strategically, it becomes clear to most people (starting in middle childhood) that open, unrestricted boasting and tattling are hardly likely to win many friends: other people will tend to see through such testimony for the reputational competition that it is. The first signs of these higher-order feedback effects are perhaps seen as early as six years of age, for although the six-year-olds in Ross and den Bak-Lammers's (1998) study tattled about as often as the four-year-olds, they engaged in positive or neutral talk about their siblings far more frequently. This suggests that by the age of six, children have already learned the value of exhibiting a positive attitude to their siblings in front of their parents. Second-order effects on the transmission of strategic information through language are likely to gain in strength through late childhood and into adolescence, as children identify more with their peer group and come to see adult authority figures as out-group members. This accounts for the inverse relationship between tattling frequency and social dominance found by Friman, Woods, et al. (2004) among teenage boys in a residential care program, where the least popular boys were perceived to do the most tattling. Further research, both experimental and observational, is needed to elaborate the precise developmental mechanisms by which ToM helps to support reputation-oriented behaviour.

## **Conclusion**

In this chapter, we have presented the hypothesis that one of the reasons why propositional theory of mind evolved in humans was to inhibit behaviour in response to the selective pressure of absent third-party punishment. Participants in our dictator game experiment (Piazza & Bering, forthcoming) used ToM to model (i) the carrying of two different pieces of information (the identity of the giver, and the amount that he or she gave) by two individuals, and (ii) the

combined effect of this information in the mind of the recipient as a result of communication between recipient and confederate. The three-to-four-year-old children in our observational study (Ingram & Bering, forthcoming) either did not possess such highly developed mind-reading abilities, or they were not yet accustomed to using these abilities strategically, to inhibit the selfish behaviour that would once have been adaptive in a world without language, in favour of the cooperative behaviour that would now be adaptive. In support of the latter interpretation, it is notable that when confronted directly with the possibility of absent third-party punishment (when an aggrieved peer said something like "I'm going to tell the teacher on you") children sometimes reversed their selfish behaviour, e.g. by handing over a toy that they had previously refused to share. Despite this, they showed little anticipation of the likelihood of tattling unless directly threatened with it. This suggests that they are missing a crucial competence which would allow them to model the spread of information within their social group – a fully developed propositional theory of mind. However, the linguistic reports of preschoolers possess important properties – truthfulness and negative bias – that would support the development of systems of indirect reciprocity and altruistic punishment. Without wanting to argue that ontogeny recapitulates phylogeny, it is tempting to speculate that early humans might at some stage during our ancestral past have made use of similar kinds of language in their reports of conspecifics' behaviour.

Recent developmental research has emphasized that language and theory of mind are inextricably linked (e.g., Astington, 2006, and chapters in Astington & Baird, 2005). We have presented one out of many possible evolutionary hypotheses which helps to explain why these two unique human competences are linked. The hypothesis that ToM evolved as an adaptive response to the threat of absent third-party punishment, inflicted on individual humans because of the spread of strategic social information via language, is no "just so story." Our hypothesis is falsifiable, and we hope that future research will attempt to disprove it. For example, neuropsychological work needs to be done to flesh out the links between the neural pathways supporting ToM and executive control and those supporting language and social cognition. If

neural pathways supporting ToM are shown to be more evolutionarily ancient than those supporting language, this would pose a major problem for our account. A similar problem would arise if animal species lacking language are shown to possess higher-order ToM skills or rudimentary systems of indirect reciprocity: there may be animals, other than humans, for which observing an individual behave antisocially towards a conspecific (which is not closely related to the observer) sometimes causes the observer to withhold cooperation from the antisocial individual. We hope that our hypothesis will inspire future research in these and other areas.

## References

- Alexander, R. D. (1987). *The biology of moral systems*. New York: Aldine de Gruyter.
- Altman, J., & Peterson, N. (1988). Rights to game and rights to cash among contemporary Australian hunter-gatherers. In T. Ingold, D. Riches & J. Woodburn (Eds.), *Hunters and gatherers*. Vol. 2. *Property, power and ideology* (pp. 75–94). Oxford: Berg.
- Astington, J. W. (2006). The developmental interdependence of theory of mind and language. In N. J. Enfield, & S. C. Levinson (Eds.), *Roots of human sociality* (pp. 179-206). Oxford: Berg.
- Astington, J. W., & Baird, J. A. (Eds.). (2005). *Why language matters for theory of mind*. Oxford: Oxford University Press.
- Baldwin, D., & Moses, L. (1994). Early understanding of referential intent and attentional focus: Evidence from language and emotion. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 133–156).
- Barclay, P., & Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B*, 274, 749-753.
- Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2(3), 412-414.
- Bering, J. M., & Shackelford, T. K. (2004). The causal role of consciousness: A conceptual addendum to human evolutionary psychology. *Review of General Psychology*, 8(4), 227-248.

- Bjorklund, D. F., Cormier, C. A., & Rosenberg, J. S. (2005). The evolution of theory of mind: Big brains, social complexity, and inhibition. In W. Schneider, R. Schumann-Hengsteler & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 147-174). Mahwah, NJ: Lawrence Erlbaum.
- Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Cambridge, MA: Harvard University Press.
- Boehm, C. (2000). Conflict and the evolution of social control. In L. D. Katz (Ed.), *Evolutionary origins of morality: Cross-disciplinary perspectives* (pp. 79-148). Thorverton: Imprint Academic.
- Bohnet, I., & Frey, B. S. (1999). The sound of silence in prisoner's dilemma and dictator games. *Journal of Economic Behavior and Organization*, 38, 43-57.
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences (USA)*, 100, 3531-3535.
- Burnham, T. C. (2003). Engineering altruism: A theoretical and experimental investigation of anonymity and gift-giving. *Journal of Economic Behavior and Organization*, 50, 133-144.
- Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72(4), 1032-1053.
- Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child development*, 69(3), 672-691.
- Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world: Inside the mind of another species*. Chicago: University of Chicago Press.

- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163-228). Oxford: Oxford University Press.
- Crystal, D. (1997). *The Cambridge encyclopedia of language* (2nd ed.). Cambridge: Cambridge University Press.
- Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation? In J. R. Krebs, & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 282-309). Oxford: Blackwell Scientific.
- Deacon, T. (1997). *The symbolic species: The co-evolution of language and the human brain*. London: Penguin.
- den Bak, Irene M., & Ross, H. S. (1996). I'm telling! The content, context, and consequences of children's tattling on their siblings. *Social Development*, 5(3), 292-309.
- Douglas-Hamilton, I., Bhalla, S., Wittemyer, G., & Vollrath, F. (2006). Behavioural reactions of elephants towards a dying and deceased matriarch. *Applied Animal Behaviour Science*, 100(1-2), 87-102.
- Dunbar, Robin I. M. (1993). Co-evolution of neocortex size, group size and language in humans. *Behavioral and Brain Sciences*, 16(4), 681-735.
- Dunbar, Robin I. M. (1999). Culture, honesty and the freerider problem. In Dunbar, Robin I. M., C. Knight & C. Power (Eds.), *The evolution of culture: An interdisciplinary view* (pp. 194-213). Edinburgh: Edinburgh University Press.
- Dunbar, Robin I. M. (2003). The social brain: Mind, language, and society in evolutionary perspective. *Annual Review of Anthropology*, 32, 163-181.
- Dunbar, Robin I. M. (2004a). Gossip in evolutionary perspective. *Review of General Psychology*, 8(2), 100-110.
- Dunbar, Robin I. M. (2004b) [1996]. *Grooming, gossip and the evolution of language* (2nd ed.). London: Faber and Faber.

- Dyer, F. C. (2002). The biology of the dance language. *Annual Review of Entomology*, 47(1), 917-949.
- Ekman, P. (1999). Basic emotions. In T. Dalgleish, & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 301-320). Chichester, Sussex: John Wiley & Sons.
- Engel, S., & Li, A. (2004). Narratives, gossip, and shared experience: How and what young children know about the lives of others. In J. M. Lucariello, J. A. Hudson, R. Fivush & P. J. Bauer (Eds.), *The development of the mediated mind* (pp. 151-174). Mahwah, NJ: Lawrence Erlbaum.
- Enquist, M., & Leimar, O. (1993). The evolution of cooperation in mobile organisms. *Animal Behaviour*, 45, 747-757.
- Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2), 63-87.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137-140.
- Fine, G. A. (1977). Social components of children's gossip. *Journal of Communication*, 27, 181-185.
- Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology*, 50, 21-45.
- Foster, E. K. (2004). Research on gossip: Taxonomy, methods, and future directions. *Review of General Psychology*, 8(2), 78-99.
- Fowler, J. H. (2005). Altruistic punishment and the origin of cooperation. *Proceedings of the National Academy of Sciences*, 102(19), 7047-7049.
- Friman, P. C., Woods, D. W., Freeman, K. A., Gilman, R., Short, M., McGrath, A. M., et al. (2004). Relationships between tattling, likeability, and social classification: A preliminary investigation of adolescents in residential care. *Behavior Modification*, 28(3), 331-348.

- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, 7(2), 77-83.
- Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26, 245-256.
- Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, 59(4), 771-785.
- Harris, P. L., & Núñez, M. (1996). Understanding of permission rules by preschool children. *Child development*, 67(4), 1572-1591.
- Hauser, M. D. (1997). *The evolution of communication*. Cambridge, MA: MIT Press.
- Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, 21(1), 101-114.
- Hockett, C. F. (1960). The origin of speech. *Scientific American*, 203, 88-96.
- Hoffman, E., McCabe, K., & Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *American Economic Review*, 86, 653-660.
- Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge: Cambridge University Press.
- Ingram, G. P. D. (2007). Tattling among pre-school children: The development of strategic social cognition. Paper delivered at the 37th annual meeting of the Jean Piaget Society, Amsterdam, June 2007.
- Ingram, Gordon P. D., & Bering, J. M. (forthcoming). *Children's tattling: A developmental precursor of gossip?* Manuscript in preparation.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *The Journal of Business*, 59(4), 285-300.
- Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum.

- Kuczaj, S., Tranel, K., Trone, M., & Hill, H. (2001). Are animals capable of deception or empathy? implications for animal consciousness and animal welfare. *Animal Welfare*, *10*(1), 161-173.
- Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology*, *4*(3), 231-244.
- Lorenz, K. (1966). *On aggression* (M. Latzke trans.). London: Methuen.
- Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford: Oxford University Press.
- Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, *393*, 573-577.
- Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*, 1291-1298.
- Núñez, M., & Harris, P. L. (1998). Psychological and deontic concepts: Separate domains or intimate connection? *Mind and Language*, *13*(2), 153-170.
- O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, *67*, 659-677.
- Perner, J. (1988). Higher-order beliefs and intentions in children's understanding of social interaction. In J. W. Astington, P. L. Harris & D. R. Olson (Eds.), *Developing theories of mind* (pp. 271-294). Cambridge: Cambridge University Press.
- Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, *3*(9), 337-344.
- Perner, J., & Wimmer, H. (1985). "John *thinks* that Mary *thinks* that...": Attribution of second-order beliefs by five- to ten-year-old children. *Journal of Experimental Child Psychology*, *39*, 437-471.
- Piazza, J. R., & Bering, J. M. (forthcoming). *Reputational concerns via gossip promote generous allocations in an economic game*. Manuscript under review.

- Povinelli, D. J., & Bering, J. M. (2002). The mentality of apes revisited. *Current Directions in Psychological Science*, 11(4), 115-119.
- Povinelli, D. J., & Eddy, T. J. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development*, 61(3).
- Povinelli, D. J., & Preuss, T. M. (1995). Theory of mind: Evolutionary history of a cognitive specialization. *Trends in Neurosciences*, 18(9), 418-424.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515-526.
- Preston, S. D., & de Waal, Frans B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25, 1-72.
- Reddy, V. (2007). Getting back to the rough ground: Deception and 'social living'. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 362(1480), 621-637.
- Richerson, P. J., Boyd, R., & Henrich, J. (2003). The cultural evolution of human cooperation. In P. Hammerstein (Ed.), *The genetic and cultural evolution of cooperation* (pp. 357-388). Cambridge, MA: MIT Press.
- Ross, H. S., & den Bak-Lammers, Irene M. (1998). Consistency and change in children's tattling on their siblings: Children's perspectives on the moral rules and procedures of family life. *Social Development*, 7(3), 275-300.
- Schoenemann, P. T. (2006). Evolution of the size and functional areas of the human brain. *Annual Review of Anthropology*, 35, 379-406.
- Semmann, D., Krambeck, H., & Milinski, M. (2004). Strategic investment in reputation. *Behavioral Ecology and Sociobiology*, 56, 248-252.

- Skinner, C. H., Cashwell, T. H., & Skinner, A. L. (2000). Increasing tootling: The effects of a peer-monitored group contingency program on students' reports of peers' prosocial behaviors. *Psychology in the Schools, 37*(3), 263-270.
- Smith, E. A., & Bliege Bird, R. (2005). Costly signaling and cooperative behaviour. In H. Gintis, S. Bowles, R. Boyd & E. Fehr (Eds.), *Moral sentiments and material interests: The foundations of cooperation in economic life* (pp. 115-148). Cambridge, MA: The MIT Press.
- Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states – the question is which ones and to what extent. *Trends in Cognitive Sciences, 7*(4), 153-156.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*(1), 35-57.
- Tyler, J. M., Feldman, R. S., & Reichert, A. (2006). The price of deceptive behavior: Disliking and lying to people who lie to us. *Journal of Experimental Social Psychology, 42*(1), 69-77.
- Wedekind, K., & Braithwaite, V. A. (2002). The long-term benefits of human generosity in indirect reciprocity. *Current Biology, 12*, 1012-1015.
- Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: The MIT Press.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655-684.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.

## Tables and Figures

	Proportion of negative reports		Proportion of true reports <sup>2</sup>	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Preschool A ( <i>n</i> = 13)	0.92	0.16	0.91	0.28
Preschool B ( <i>n</i> = 21)	0.94	0.08	0.88	0.23
2-year-old on 4-year-old siblings	0.87	<i>N/A</i>	0.96	0.13
4-year-old on 2-year-old siblings	0.80	<i>N/A</i>	0.91	0.19
4-year-old on 6-year-old siblings	0.70	<i>N/A</i>	0.96	0.08
6-year-old on 4-year-old siblings	0.56	<i>N/A</i>	0.94	0.13

*Table 1 – Comparison of tattling between siblings and in the two preschools*

	No threat of gossip		Threat of gossip	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
No reputational concern	4.22	1.73	4.67	1.09
Reputational concern	4.17	2.23	5.44	1.15

*Table 2 – Mean allocation scores (and standard deviations) for each condition (*n* = 18) of dictator game*

<sup>2</sup> Truth values are given for negative reports only, since Ross and den Bak-Lammers (1998) do not report the percentages of non-negative reports that were true.