

THE MORAL DIMENSIONS OF HUMAN SOCIAL INTELLIGENCE

Domain-specific and Domain-general Mechanisms

Valerie Stone

Human moral behaviour ranges from vicious cruelty to deep compassion, and any explanation of morality must address how our species is capable of such a range. Darwin argued that any social animal, with sufficient intellectual capacity, would develop morality. In agreement, I argue that human morality is unique in the animal kingdom not because of any particular moral capacity, but because some very abstract cognitive abilities that are unique to our species are layered on top of phylogenetically older emotional instincts for aggression and for empathy. I review research on several components of empathy, intersubjectivity and theory of mind, detailing which of these capacities is uniquely human, and highlighting relevant neuroanatomy for each ability. Finally, I describe our abstract cognitive abilities for recursion and metarepresentation, and argue that these are uniquely human. When these abilities interact with our older social abilities (empathy, intersubjectivity) we are able to reason abstractly about others' mental states and how to affect them. Thus, it is these abstract cognitive capacities that give us the ability to be both cruel and compassionate, but it is our ancient ability for empathy that keeps us moral. Morality is, in a very real sense, a gift from our ancestors.

The mind is a complex thing... in the sense that it is a product of ancient reflexive and emotional components, overlain by a veneer of reason... we have not entirely emancipated ourselves from the brain structures that governed the behavior of some very remote ancestors indeed. And it is precisely this interaction of the ancient with the new that makes us not only unique in many very admirable ways, but also uniquely dangerous—as much to ourselves as to the rest of the living world... [N]ature, while having placed a unique and potentially highly destructive capacity in the hands of *Homo sapiens*, is under no obligation to ensure we use that capacity wisely. (Tattersall 1998, 234)

Introduction

Humans are an odd combination of emotion and reason. Our moral psychology reflects this twin nature. To understand human morality, consider its extremes: compassion and cruelty. Compassion is deep concern for the well-being of others, understanding a situation from another's point of view, and undertaking carefully planned actions to help.

Cruelty involves careful modelling of others' mental states to take deliberate, painful action to cause harm. Both extremes involve using abstract, high-level cognitive abilities as well as older emotional systems. Without those high-level abilities, these extremes of morality would not be possible.

I will argue that human morality is unique in the animal kingdom not because humans have evolved a 'moral instinct'—but because some very abstract cognitive abilities that are unique to our species are layered on top of phylogenetically older social capacities and emotions.

One such ability is language, which gives us the capacity for symbolic representation and abstract syntax. Underlying language syntax is the still more abstract capacity for recursion, the processing of embedded representations (Corballis 2003). Another high-level ability is 'theory of mind', the capacity to infer the explicit content of others' mental states. Finally, 'executive function' is a term used to refer broadly to humans' capacity to flexibly control attention and inhibit action, to hold several pieces of information in mind while performing operations on the information (known as working memory), and to plan complex sequences of action. Using these high-level and abstract capacities, we can engage in a wide range of moral judgements and behaviours.

Evolutionary psychology argues that much of human behaviour, including morality, is produced by domain-specific cognitive systems, meaning systems operating on particular types of input, using certain types of representations, and performing certain types of computations on those representations (Buss and Kenrick 1998; Tooby and Cosmides 1992). Domain-specific cognitive systems are *adaptations*, specialized for particular functions and designed by natural selection over millions of generations to have information-processing features specific to the cognitive problem being solved (Cosmides and Tooby 1987). Because adaptations are complex, each design feature must evolve layered onto previous features; special design in an adaptation thus takes a long evolutionary time to emerge (Dawkins 1987; Williams 1966).

It is true that many aspects of our emotional life and social motivations are shared in common with other primates and mammals who faced the same adaptive problems. There have been tens or hundreds of millions of years to refine neural adaptations for those domains of behaviour. Abilities that are recent and unique to the hominid line, however, are not likely to depend solely on domain-specific systems, because there simply has not been enough time for such specialized design to evolve (Stone 2003, 2005, forthcoming). Thus, to understand how our social intelligence serves human morality, we must look at how domain-specific systems that we share with other primates interact with more general cognitive abilities that are uniquely human.

Language serves many social functions. Before children learn to use symbols, they are often bound to their current perceptual environment, and have difficulty inhibiting behaviour that is strongly dictated by their environment (Kirkham, Cruess, and Diamond 2003). Symbols, however, provide cognitive distance from immediate circumstances. Once children can use language, they get much better at resisting environmental demands (cf. Deak, Ray, and Breneman 2003; Kirkham, Cruess, and Diamond 2003; Müller et al. 2004). Thus, the simple fact of having a language capacity means that we can make moral choices that involve using that kind of cognitive distancing. Nevertheless, though language can have social functions, it involves many cognitive skills that have nothing to do with socially in particular: segregation of an auditory stream into units, symbolic representation, metarepresentation, recursion and syntax (cf. Corballis 2003; Hauser, Chomsky, and Fitch

2002). These computational aspects of language solve information-processing problems that do not concern social information.

Executive function, i.e. flexible control of attention and action, working memory, and planning complex sequences of action, is also useful in social behaviour. Social interactions constantly change over time, requiring flexible control of attention to track the most important events as they unfold. Keeping track of a conversation, or of how others' reactions are changing over time, requires working memory. Selecting which social actions to perform out of many possible options means inhibiting some actions in favour of others, requiring executive function. And finally, engaging in any kind of long-term social strategy, such as courtship or politics, requires planning and sequencing abilities. However, like language, executive function is not specifically social. It is a more general set of abilities, solving the problem of coordinating the output of many domain-specific systems to produce behaviour.

Of the uniquely human abstract cognitive abilities listed above, theory of mind (ToM) is the only *specifically social cognitive ability*, and thus may be a key factor in moral choices. I would like to discuss this cognitive ability in more detail to explore the architecture of social and moral cognition. It depends on specific cognitive precursors. Collectively, these give us *intersubjectivity*, an implicit understanding of others' goals, intentions, attentional focus, emotions and desires. Young children and great apes have intersubjectivity (for reviews, see Stone 2005; Suddendorf and Whiten 2001). However, there is no convincing evidence yet that they can make complex ToM inferences about the explicit contents of others' mental states (Suddendorf and Whiten 2003). ToM is layered on top of intersubjectivity and its development depends on the forging of connections between brain circuits for domain-general cognition and circuits specialized for aspects of intersubjectivity. ToM's uniquely human components may not be specifically social (Suddendorf 1999; Stone 2005). This review focuses on when various abilities emerge, which abilities seem to be domain-specific, and how domain-general abilities unique to humans might build on domain-specific ones shared with primates to produce complex social cognition, and thus, moral cognition.

Components of Intersubjectivity

Animacy Detection

ToM inferences are made only about animate entities; therefore one of the fundamental components of any kind of social perception is the detection of animacy, that is, perceiving entities in the world as having self-propelled, goal-directed motion. Animacy detection emerges early in development, and is shared with other animals. Infants as young as 16–20 weeks visually distinguish animate from inanimate motion (Crichton and Lange-Küttner 1999). In macaque monkeys, both unexpected touch and the sight of others' hand movements (but not the monkey's own) activate populations of cells in the superior temporal sulcus (STS) (Perrett et al. 1990).

Neuroimaging in humans points to the STS as well as other areas specialized for animacy detection (Blakemore et al. 2003). Inanimate motion primarily activated different areas than animate motion. The brain processes animate objects differently from inanimate ones. Animacy detection appears to be an ancient and domain-specific component of human social perception.

Inferring Goals and Intentions

Infants from very early on begin to distinguish actions that are intentional, and to discern an actor's goal. Infants between five and nine months can differentiate accidental from intentional behaviour (Woodward 1999), and by 15 months, classify an action according to the action's goal (Sibra et al. 2003). These results show an implicit understanding of intentions and goals. Chimpanzees and orang-utans can also distinguish visually between accidental and intentional actions (Call and Tomasello 1998). Monkeys do not seem to make this distinction, putting the date for this ability in a common ancestor about 14 million years ago, assuming homology (Wildman et al. 2003). The brain circuitry involved appears to be the STS, and appears to be domain-specific, with cells that respond to the combination of gaze and reaching not responding to gaze or reaching alone (Jellema et al. 2000).

Eye Gaze Detection and Gaze Following

Newborn infants are sensitive to direct eye contact versus averted gaze at two to five days after birth, and electrophysiological measures of four-month-old infants show distinct brain activity patterns to direct versus averted gaze (Farroni et al. 2002, 2003). The temporal lobes seem to be crucial for gaze direction detection. In macaques, cells in the STS respond differentially to certain directions of eye gaze and head angle, and lesions in STS impair differentiation of gaze direction (Campbell et al. 1990; Perrett et al. 1990). In human adults, there is evidence from neuroimaging studies that the STS and amygdala are activated specifically in tasks involving detecting direction of gaze (Hoffman and Haxby 2000; Kawashima et al. 1999). Young et al. (1995) also found that a patient with bilateral amygdala lesions had difficulty correctly detecting direct eye contact. The brain systems involved in eye direction detection in humans appear to be similar to those in monkeys, and thus may be shared through common descent. These systems respond to images of eyes or a head looking in a certain direction, but not to visually similar, non-social stimuli (Campbell et al. 1990). The second step after gaze direction detection is the ability to refocus one's attention where another person is looking, known as 'gaze following'. Infants follow gaze by the age of three or four months. Primates are clearly sensitive to eye gaze and eye contact, with many species of monkeys able to detect and follow gaze direction (Anderson and Mitchell 1999; Ferrari et al. 2000; but see Itakura 1996 for conflicting results on rhesus macaques; Okamoto et al. 2002; Tomasello, Hare, and Fogelman 2001; Vick and Anderson 2003). Gaze detection and following appear to be domain-specific and ancient.

Gaze Monitoring

Young children do more with gaze direction than just follow it. They gradually begin to respond as if it means something about others' mental states. Between nine and 24 months of age, children begin not just to follow gaze reflexively, but to actively monitor others' gaze direction as a source of information (Buttenworth and Jarrett 1991; Phillips, Baron-Cohen, and Rutter 1992). Children between nine and 18 months old will look at an adult's eyes when presented with an ambiguous situation, indicating an implicit understanding that eye gaze direction can provide information about the world (Phillips, Baron-Cohen, and Rutter 1992). Children aged 14–19 months can learn novel words for novel objects only if they could rely on the speaker's direction of gaze to find the correct

referent for the word (Baldwin 1993). From roughly nine to 19 months, infants enter this new stage of 'gaze monitoring', actively seeking information about others' gaze direction instead of merely following it when presented with the visual stimulus of another's gaze. Apes and some monkeys can do this as well (Hare et al. 2000; Kumashiro et al. 2002; Suddendorf and Whiten 2001, 2003). Causes of ambiguous behaviour are sought in the social rather than the physical world; thus gaze monitoring also seems to be domain-specific (Stone forthcoming).

Joint Attention and Pretend Play

'Joint attention' also emerges between 18 and 24 months, and means that the child calls adults' attention to particular objects by pointing or holding up something for them to see. Children at this stage become active in trying to affect others' attention (Baron-Cohen 1995; Franco and Buttenworth 1996). Joint attention also requires a secondary representation that is de-coupled from the child's primary perceptual representation of reality. The child integrates information about whether an adult is paying attention (binary yes/no), the rough spatial location of objects, and a repertoire of actions that generally succeed in engaging adults' attention (holding things up, pointing). Thus, what may have emerged at the gaze-monitoring and joint-attention stage is not only a new domain-specific ability, but also a new domain-general ability, the capacity to form secondary representations. Pretend play also emerges at this age, and also appears to depend crucially on secondary representations (Suddendorf and Whiten 2001). Secondary representations can be used for many tasks besides these social tasks, e.g., means-end reasoning, or finding hidden objects (Suddendorf and Whiten 2001). This domain-general ability can take the output of domain-specific systems for gaze direction and goal detection, and integrate them with the child's own actions.

Apes and monkeys have been found to use gaze monitoring and joint attention (Byrne 2001; Hare et al. 2000; Kumashiro et al. 2002, 2003; Povinelli et al. 2003; for a review, see Stone forthcoming). These abilities are at least 14 million years old if only shared with apes, and possibly 26 million years old if also shared with monkeys. I know of no study to date that focuses specifically on systems that might mediate joint attention or pretend play in the brain.

Developmental psychologists debate whether pretence involves representing others' mental states or is simply treated as a special kind of action (Leslie 1987; Wellman and Lagattuta 2000). Neuroscience research could address this debate. If children treat pretence as a special kind of action, areas involved in pretence might overlap with areas involved in representing actions not currently being done, e.g., supplementary motor area. If pretence requires children to represent mental states, then the same areas active for full metarepresentational ToM should be active for pretence. Although one functional magnetic resonance imaging (fMRI) study of adult pretence showed activation in ToM areas (German et al. forthcoming), this issue in children remains unresolved.

Empathy

Much research on how human children develop the capacity for moral choices has focused on empathy and pro-social behaviour (helping, comforting), the building blocks of altruism (Preston and deWaal 2002). Newborn infants will cry when other infants cry

(Sagi and Hoffman 1976; Sinner 1971). Around the first year, children also begin to look for others' emotional reactions to ambiguous situations, and respond accordingly—if the mother seems calm in a new situation, the child is more likely to remain calm (Boccia and Campos 1989). Children who are a bit older, between 15 and 24 months, go further in their reactions to someone else being upset—they actively try to help. They might be physically affectionate or might bring a comforting object to the person, even if it is not what the other person might find most comforting. With language developing around the same time, toddlers will also use words to comfort, 'Don't cry, Mommy'. Children at this age seem to have the goal of relieving others' distress, because they will keep trying different strategies until one works (Zahn-Waxler and Radke-Yarrow 1982; Zahn-Waxler, Radke-Yarrow, and King 1979).

Many animals, even rats, will show the most basic form of empathy: being upset by seeing another member of their species in distress (Preston and deWaal 2002). Monkeys and apes will go further than merely being upset by another's distress, and will try to reduce the other's upset. Our common ancestor with monkeys and apes probably had this capacity as well. Chimpanzees actively comfort other chimpanzees who are distressed. When one chimpanzee has been the target of aggression, group members may approach to console him or her, embracing or grooming the 'victim' of aggression (Flack and deWaal 2000). Third-party hugging and consolation seem to happen more often after a conflict than at other times; chimpanzees truly do seem to show more concern when an individual is in distress than at other times. They also show more positive behaviour towards the victim of aggression than towards the perpetrator (deWaal and Aurell 1996). The pro-social behaviour that follows from empathy, attempting to relieve another's distress, seems to be something our primate forebears have handed down to us.

Empathy may involve many brain regions. First are areas involved in recognizing others' emotional facial expressions, typically right hemisphere more than left, somatosensory areas in parietal cortex that are involved in sensing our own facial movements, and the amygdala (Adolphs et al. 1994, 2000; Etcoff 1984; Young et al. 1995). Our ability to recognize others' emotions appears to be intimately tied to the same brain areas involved in our own experience and expression of emotion. Mirror neurons—cells active to the same degree whether perceiving one's own or another's movement—have also been proposed as crucial for empathy (Gallese 2003). Empathy studies using fMRI implicate extensive circuitry involved in recognizing others' emotions, perceiving others' pain and inferring emotions from verbal scenarios: motor planning areas, superior temporal lobes, amygdala, insula, anterior cingulate, orbitofrontal cortex (e.g., Carr et al. 2003; Farrow et al. 2001; Jackson, Meltzoff, and Decety 2005). Patient studies have tended to point more consistently to the amygdala and orbitofrontal regions, particularly right hemisphere (Perry et al. 2001; Shamay-Tsoory et al. 2005; Stone et al. 2003; Young et al. 1995). A clear model of the brain basis of the tendency to feel distress when others feel distress—the definition of empathy—awaits empirical work that operationalizes this definition closely.

Understanding Others' Desires

By the age of two, children seem to implicitly understand others' emotions and focus of attention. Beginning around age two, children elaborate their emotional knowledge and begin to use language about desire, 'I want', 'she likes' (Wellman, Cross, and Watson 2001; Wellman and Lagattuta 2000). They seem to understand that people's attitudes and

emotions towards various objects can be used to predict what they will do (Wellman and Lagattuta 2000) and that other people can have different desires to the child's own (Repacholi and Gopnik 1997; Wellman and Liu 2004; Wellman and Wooley 1990).

Desire can change without external reality changing. Thus, this developmental step represents the first stage at which children display understanding of something that is uniquely *mental*, private, unobservable and changeable. This implicit understanding of the private and changeable nature of mental things, which could be called 'mentalism', is a genuinely new development in ToM, and seems specific to ToM, not explainable by reference to perceptually based cognitive abilities. Understanding desire thus may set the stage for the next step in ToM development at age three (Wellman 1990; Wellman and Liu 2004).

Theory of Mind Proper: Explicit Understanding of Others' Mental States

Children go beyond intersubjectivity to a full representational model of others' minds when they develop ToM: the ability to understand that others have beliefs about the world which may be incorrect, and to reason about those beliefs using forms of representation which are not tied to a particular perceptual modality. Thus, testing whether a subject can understand that someone else holds a false belief has long been held to be the key test of ToM¹ (Dennett 1978). Children usually pass this test some time between the third and fourth years of life, but it is rare that three-year-olds can pass it (Wellman, Cross, and Watson 2001; Wellman and Liu 2004).

The key to ToM is the emergence of 'metarepresentation', the ability to explicitly represent representations or *representations* (Baron-Cohen 1995; Leslie 1994; Perner 1991). It is metarepresentation that enables children to form beliefs about the beliefs of others.

However, many other cognitive abilities also contribute to being able to pass an explicit false belief task. In particular, solving false belief tasks depends on executive control, being able to inhibit the inappropriate response—the subject's knowledge of the true state of reality—in order to answer with the perhaps less salient correct response—the other person's mental state (Carlson and Moses 2001; Carlson, Moses, and Klaxton 2004; Flynn, O'Malley, and Wood 2004). Children can pass false belief tasks slightly earlier if the task demands are changed so that less inhibitory control is required, e.g., making the current state of reality less salient (Wellman and Lagattuta 2000). False belief tasks also depend on working memory and sequencing, as the subject has to keep in mind all the elements of the story as it unfolds in order, and how those elements are changing with respect to each other (Keenan 1998; Stone, Baron-Cohen, and Knight 1998). Thus, someone who has deficits in inhibition or working memory could easily fail a false belief task while having intact metarepresentational abilities.

Furthermore, metarepresentation may be just one example of a more general cognitive ability: recursion. If metarepresentation is simply one type of recursion rather than a separate ability, difficulties with recursion could cause failures on the false belief task (Corballis 2003). Whether separate abilities or related, recursion and metarepresentation are both domain-general cognitive abilities, not limited to social cognition, that interact with intersubjectivity and mentalism to produce what we call explicit ToM (Corballis 2003; De Villiers 2000; Suddendorf 1999). In development, the ability to use and understand sentence complements and embedded relative clauses (recursive syntactic

structures) precedes the ability to pass false belief tests (De Villiers 2000; De Villiers and Pyers 2002; Smith, Apperly, and White 2003).

The idea that explicit ToM is dependent on the metarepresentational competence needed for such complex grammatical structures is consistent with results on the cognitive abilities of apes. Chimpanzees, given an explicit false belief test in which they had to choose a container with food, either one that was correctly marked or one in which an experimenter had a false belief about the validity of the mark, consistently failed the task (Call and Tomasello 1999; Hare et al. 2001). They seem to lack metarepresentation (Suddendorf 1999; Suddendorf and Whiten 2003). Similarly, chimpanzees who have been taught to use signs and symbols to refer to things never ask for the names of things, showing that their understanding of symbols may be more associative than representational; they have never been observed to use complex syntax at all, much less sentence complements (Pinker 1994; Snowdon 2001). Apes also do not show any evidence of either episodic memory or future planning (Suddendorf 1999; Suddendorf and Busby 2003). Thus, metarepresentation and recursion may be uniquely human capacities (Corballis 2003; Suddendorf 1999).

The most phylogenetically and developmentally advanced *domain-specific* aspect of ToM is the capacity for intersubjectivity we share with toddlers and great apes. The evolutionary breakthrough for humans was the ability to link metarepresentation/recursion and executive function to intersubjectivity (Stone and Gerrans forthcoming).

ToM in the Brain

Social neuroscience has been studying ToM only for a decade, and thus neuroscience research on ToM is still very much in its infancy. Much ToM research in neuroscience has not been done with proper controls for non-social metarepresentation, working memory, inhibitory demands of tasks, or other executive functions, nor has it been done with a clear definition of which types of mental states (e.g., intention, belief, desire) are being tapped by various tasks (Apperly, Samson, and Humphreys forthcoming; Stone 2005). I include my own research in this criticism. The body of research in this area claims variously that ToM might be processed in superior temporal areas, temporal pole, the amygdala, temporal-parietal junction, medial frontal cortex, orbitofrontal cortex, and/or frontal pole (reviewed in Apperly, Samson, and Humphreys forthcoming; Stone 2005). I argue that so many candidate brain areas have emerged as important for ToM because different areas subserve different aspects of ToM, some subserving domain-general functions, some domain-specific (Stone 2005). Indeed, it appears from patient and autism research that there are no cases of impaired ToM without accompanying deficits in either intersubjectivity, or domain-general abilities such as executive function or metarepresentation (Apperly et al. 2004; Apperly, Samson, and Humphreys 2005; Stone and Gerrans 2006).

Domain-specificity, Domain-generality and Morality

Darwin thought that any social animal would develop a similar moral sense to humans if it had intellectual capacities equal to those of humans: 'any animal whatever, endowed with well marked social instincts, ... would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well, or nearly as well developed, as in man [*sic*]' (Darwin 1871, 471–72). Darwin distinguished intellectual capacities from instincts and argued that our moral sense depends on those intellectual capacities.

Following this review of intersubjectivity and theory of mind, I find myself much in agreement. The emotional and social aspects of morality depend on ancient systems for intersubjectivity that seem to be shared with our primate cousins, particularly the great apes. However, layered on top of this, the intellectual capacities that distinguish our species give humans a psychology that involves complex inferences about others' internal states, symbolic representation of and communication about moral standards, and planning moral actions in terms of future goals. Metarepresentation, recursion, language and executive function all interact with our more ancient capacity for intersubjectivity, and a thorough understanding of human moral psychology must tease apart these different cognitive systems.

I disagree with Darwin only in his emphasis. I would say that our ancient social instincts are the driving force of morality, while our intellect merely expands the range of behaviours that can be driven by this force. The extreme poles of morality—compassion and cruelty—are possible because we have high-level abilities for planning, inhibition, understanding others' mental states, and language. We use language not just to communicate information, but to construct social narratives about who is a member of the in-group, who is the enemy, and which people and actions should be considered moral. This symbolic representation of the social world influences our moral behaviour, but does not itself provide the motivational force. Cruelty is what happens when we use our symbolic capacity to define another person as 'the enemy', and use theory of mind and executive function to plan an outlet for our ancient instincts for aggression. Compassion is what happens when we use our symbolic capacity to define another person as part of our in-group, and use our theory of mind and executive function to plan ways to benefit them, using ancient instincts for empathy. Although abstract domain-general cognition makes the range of human moral behaviour possible, this cognition is not the engine of morality. Rather, our emotional capacities, empathy on the positive side, and aggression on the negative side, provide the motivation for human moral behaviour. Feeling for another's suffering, wanting to help relieve that suffering, reconciling after conflict, assessing intentions—all these capacities are aspects of mind that have been handed down to us from our common ancestor with other primates. Combined with our more complex cognition, they make us a moral animal. Morality is, in a very real sense, a gift from our ancestors.

NOTE

- Two basic kinds of false belief tasks have been used with children, *location change* tasks and *unexpected contents* tasks. In a location change task, the subject is told a short story (and shown pictures to go with the story), or the story is acted out with toy figures) in which character A puts an object in location 1, and then turns away or goes out of the room. Character B moves the object to location 2 while A cannot see (or, if the object is animate, like a pet, it moves itself), and then the subject is asked where A will look for the object, location 1 or location 2.

REFERENCES

- ADOURPHIS, R., H. DAMASIO, D. TRANEL, G. COOPER, and A. R. DAMASIO. 2000. A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience* 20 (7): 2683–90.

- ADOLPHS, R., D. TRANEL, H. DAMASIO, and A. R. DAMASIO. 1994. Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* 372: 669–72.
- ANDERSON, J., and R. MITCHELL. 1999. Macaques but not lemurs co-orient visually with humans. *Folia Primatologica* 70 (1): 17–22.
- APPLEY, I. A., D. SAMSON, C. CHIAVARINO, and G. W. HUMPHREYS. 2004. Frontal and temporo-parietal lobe contributions to Theory of Mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience* 16 (10): 1773–84.
- APPLEY, I. A., D. SAMSON, and G. W. HUMPHREYS. 2005. Domain-specificity and theory of mind: Evaluating neuropsychological evidence. *Trends in Cognitive Science* 9 (12): 572–77.
- BALDWIN, D. 1993. Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language* 20 (2): 395–418.
- BARON-COHEN, S. 1995. *Mindblindness: An essay on autism and theory of mind*. Cambridge, Mass.: MIT Press.
- BLAKEMORE, S. J., P. BOYER, M. PACHOT-CLOUGHARD, A. MELTZOFF, C. SEGEBARTH, and J. DECEY. 2003. The detection of contingency and animacy from simple animations in the human brain. *Cerebral Cortex* 13 (8): 837–44.
- BOCCIA, M., and J. J. CAMPOS. 1989. Maternal emotional signals, social referencing, and infants' reactions to strangers. *New Directions for Child Development* 44 (Summer): 25–49.
- BUSS, D., and D. T. KENRICK. 1998. Evolutionary social psychology. *The Handbook of Social Psychology (4th Edition)* 2: 982–1026.
- BUTTERWORTH, G., and N. JARRETT. 1991. What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology* 9 (1): 55–72.
- BYRNE, R. W. 2001. Social and technical forms of primate intelligence. In *Tree of origin: What primate behavior can tell us about human social evolution*. Cambridge, MA: Harvard University Press.
- CALL, J., and M. TOMASELLO. 1998. Distinguishing intentional from accidental actions in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *Journal of Comparative Psychology* 112: 192–206.
- . 1999. A nonverbal false belief task: The performance of children and great apes. *Child Development* 70 (2): 381–95.
- CAMPBELL, R., C. HEYWOOD, A. COWEY, M. REGARD, and T. LANDIS. 1990. Sensitivity to eye gaze in prosopagnosic patients and monkeys with superior temporal sulcus ablation. *Neuropsychologia* 28 (11): 1123–42.
- CARLSON, S., and L. MOSES. 2001. Individual differences in inhibitory control and children's theory of mind. *Child Development* 72 (4): 1032–53.
- CARLSON, S., L. MOSES, and L. KILPATRICK. 2004. Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology* 87 (4): 299–319.
- CARR, L., M. IACOBONI, M. C. DUBEAU, J. C. MAZZIOTTA, and G. L. LENZ. 2003. Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences USA* 100 (9): 5497–502.
- CORBALLIS, M. 2003. Recursion as the key to the human mind. In *From mating to mentality: Evaluating evolutionary psychology*, edited by K. Sterelny and J. Fitness. New York: Psychology Press.
- COSMIDES, L., and J. TOOBY. 1987. From evolution to behavior: Evolutionary psychology as the missing link. In *The latest on the best: Essays on evolution and optimality*, edited by J. Dupre. Cambridge, Mass.: MIT Press.
- CRICHTON, M., and C. LANGE-KUTNER. 1999. Animacy and propulsion in infancy: Tracking, waving, and reaching to self-propelled and induced moving objects. *Developmental Science* 2 (3): 318–24.
- CIRBA, G., S. BIRN, O. KOOS, and G. GERGELY. 2003. One-year-old infants use teleological representations of actions productively. *Cognitive Science* 27 (1): 111–33.
- DARWIN, C. 1871. *The descent of man and selection in relation to sex*. New York: Modern Library.
- DAWKINS, R. 1987. *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. New York: W. W. Norton and Company.
- DE VILLIERS, J. 2000. Language and theory of mind: What are the developmental relationships? In *Understanding other minds: Perspectives from developmental cognitive neuroscience*, 2nd ed., edited by S. Baron-Cohen, H. Tager-Flusberg, and D. Cohen. Oxford: Oxford University Press.
- DE VILLIERS, J., and J. PETERS. 2002. Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief understanding. *Cognitive Development* 17: 1037–60.
- DEWAAL, F., and F. AURELL. 1996. *Consolation, reconciliation and a possible cognitive difference between macaques and chimpanzees*. Cambridge: Cambridge University Press.
- DEK, G. O., S. D. RAY, and K. BRENNEMAN. 2003. Children's perseverative appearance–reality errors are related to emerging language skills. *Child Development* 74 (3): 944–64.
- DENNETT, D. 1978. Beliefs about beliefs. *Behavior and Brain Sciences* 4: 568–70.
- ETCOFF, N. L. 1984. Perceptual and conceptual organization of facial emotions: Hemispheric differences. *Brain and Cognition* 3: 385–412.
- FARRON, T., G. CIRBA, F. SIMON, and M. JOHNSON. 2002. Eye contact detection in humans from birth. *Proceedings of the National Academy of Science USA* 99 (14): 9602–5.
- FARRON, T., E. MAUSFIELD, C. LAI, and M. JOHNSON. 2003. Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *Journal of Experimental Child Psychology* 85: 199–212.
- FARROW, T. F., Y. ZHENG, I. D. WILKINSON, S. A. SPENCE, J. F. DEAKIN, N. TARRIER, et al. 2001. Investigating the functional anatomy of empathy and forgiveness. *Neuroreport* 12 (11): 2433–38.
- FERRARI, P. F., E. KOHLER, L. FOGASSI, and V. GALLESSE. 2000. The ability to follow eye gaze and its emergence during development in macaque monkeys. *Proceedings of the National Academy of Science USA* 97 (25): 13997–14002.
- FLAGG, J. C., and F. B. M. DEWAAL. 2000. 'Any animal whatever': Darwinian building blocks of morality in monkeys and apes. *Journal of Consciousness Studies* 7 (1/2): 1–29.
- FLYNN, E. C. O'MALLEY, and D. WOOD. 2004. A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science* 7 (1): 103–15.
- FRANCO, F., and G. BUTTERWORTH. 1996. Pointing and social awareness: Declaring and requesting in the second year. *Journal of Child Language* 23 (2): 307–36.
- GALLESSE, V. 2003. The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* 36 (4): 171–80.
- GERMAN, T. J., L. NIENHAUS, M. P. ROHARTY, B. GIESBRECHT, and M. B. MILLER. 2004. Neuro correlates of detecting pretense: Automatic engagement of the intentional stance under covert conditions. *Journal of Cognitive Neuroscience* 16 (10): 1805–17.
- HABE, B., J. CALL, B. AGNETTA, and M. TOMASELLO. 2000. Chimpanzees know what conspecifics do and do not see. *Animal Behaviour* 59: 771–85.

- HARE, B., J. CALL, and M. TOMASELLO. 2001. Do chimpanzees know what conspecifics know? *Animal Behaviour* 61: 139–51.
- HAUSER, M. D., N. CHOMSKY, and W. T. FITCH. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569–79.
- HOFFMAN, E. A., and J. V. HAXBY. 2000. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience* 3 (1): 80–84.
- TAJIMA, S. 1996. An exploratory study of gaze-monitoring in non-human primates. *Japanese Psychological Research* 38: 174–80.
- JACKSON, P. L., A. N. MELTZOFF, and J. DECEY. 2005. How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage* 24 (3): 771–79.
- JELLEM, T., C. I. BAKER, B. WICKER, and D. I. PERRETT. 2000. Neural representation for the perception of the intentionality of actions. *Brain and Cognition* 44 (2): 280–302.
- KAWASHIMA, R., M. SUGIURA, T. KATO, A. NAKAMURA, K. HATANO, K. ITO, et al. 1999. The human amygdala plays an important role in gaze monitoring—A PET study. *Brain* 122: 779–83.
- KEMAN, T. 1998. Memory span as a predictor of false belief understanding. *New Zealand Journal of Psychology* 27 (2): 36–43.
- KIRKHAM, N. Z., L. CRUICK, and A. DIAMOND. 2003. Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science* 6 (5): 449–76.
- KUMASHIRO, M., H. ISHIVASHI, S. TAKIWA, and A. IRIKI. 2002. Bidirectional communication between a Japanese monkey and a human through eye gaze and pointing. *Current Psychology of Cognition* 21 (1): 3–32.
- KUMASHIRO, M., H. ISHIVASHI, Y. UCHINAMA, S. TAKIWA, A. MURATA, and A. IRIKI. 2003. Natural imitation induced by joint attention in Japanese monkeys. *International Journal of Psychophysiology* 50: 81–99.
- LESLIE, A. M. 1987. Pretence and representation: The origins of 'theory of mind'. *Psychological Review* 94: 412–26.
- . 1994. Pretending and believing: Issues in the theory of TOMM. *Cognition* 50: 211–38.
- MÜLLER, U., P. D. ZELAZO, S. HOOD, T. LEONE, and L. ROHNER. 2004. Interference control in a new rule use task: Age-related changes, labeling, and attention. *Child Development* 75 (5): 1594–609.
- OKAMOTO, S., M. TOMONAGA, K. ISHII, N. KAWAI, M. TANAKA, and T. MATSUZAWA. 2002. An infant chimpanzee (*Pan troglodytes*) follows human gaze. *Animal Cognition* 5 (2): 107–14.
- PERNER, J. 1991. *Understanding the representational mind*. Cambridge, Mass.: MIT Press.
- PERRETT, D., M. HARRIES, A. MISTLIN, J. HEITANEN, P. BENSON, R. BEVAN, et al. 1990. Social signals analyzed at the single cell level: Someone is looking at me, something touched me, something moved. *International Journal of Comparative Psychology* 4 (1): 25–55.
- PERRY, R. J., H. R. ROSEN, J. H. KRAMER, J. S. BEER, R. L. LEVENSON, and B. L. MILLER. 2001. Hemispheric dominance for emotions, empathy and social behaviour. *Neurocase* 7 (2): 145–60.
- PHILLIPS, W., S. BARON-COHEN, and M. RUTTER. 1992. The role of eye-contact in the detection of goals: Evidence from normal toddlers, and children with autism or mental handicap. *Development and Psychopathology* 4: 375–83.
- PINKER, S. 1994. *The language instinct*. New York: William Morrow.
- POVINELLI, D. J., L. A. THEALL, J. E. REAVY, and S. DUNPHY-LELLI. 2003. Chimpanzees spontaneously alter the location of their gestures to match the attentional orientation of others. *Animal Behaviour* 66 (1): 71–79.
- PRESTON, S. D., and F. B. M. DEWAAL. 2002. Empathy: Its proximate and ultimate bases. *Behavioral and Brain Sciences* 25: 1–72.
- REPACHOU, B. M., and A. GOPNIK. 1997. Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology* 33 (1): 12–21.
- SAGI, A., and M. L. HOFFMAN. 1976. Empathic distress in humans. *Developmental Psychology* 12: 175–76.
- SHAMAY-TSOORY, S. G., R. TOMER, B. D. BERGER, D. GOUDSHER, and J. AHARON-PERETZ. 2005. Impaired "affective theory of mind" is associated with right ventromedial prefrontal damage. *Cognitive and Behavioral Neurology* 18 (1): 55–67.
- SIMMER, M. L. 1971. Newborn responses to the cry of another infant. *Developmental Psychology* 5: 136–50.
- SMITH, M., I. APPRELY, and V. WHITE. 2003. False belief reasoning and the acquisition of relative clause sentences. *Child Development* 74 (6): 1709–19.
- SNOWDON, C. T. 2001. From primate communication to human language. In *Tree of origin: What primate behavior can tell us about human social evolution*, edited by F. deWaal. Cambridge, Mass.: Harvard University Press.
- STONE, V. E. 2003. Foolproof and fossil-free no more: Evolutionary psychology needs archaeology. *Behavioral and Brain Sciences* 25 (3): 420–21.
- . 2005. Theory of mind and the evolution of social intelligence. In *Social neuroscience: People thinking about people*, edited by J. Gaciopio. Cambridge, Mass.: MIT Press.
- In press. An evolutionary perspective on domain-specificity and social intelligence. In *Social cognitive neuroscience*, edited by P. Winkielman and E. Harmon-Jones. New York: Guilford.
- STONE, V. E., S. BARON-COHEN, A. C. CALDER, J. KEANE, and A. W. YOUNG. 2003. Acquired theory of mind impairments in individuals with bilateral amygdala lesions. *Neuropsychologia* 41: 209–20.
- STONE, V. E., S. BARON-COHEN, and R. T. KNIGHT. 1998. Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience* 10: 640–56.
- STONE, V. E. and P. GERRANS. 2006. Does the normal brain have a theory of mind? *Trends in Cognitive Science* 10 (1): 3–4.
- SUDENDORF, T. 1999. The rise of the metacognition. In *The descent of mind: Psychological perspectives on hominid evolution*, edited by M. C. Corballis and S. Lea. London: Oxford University Press.
- SUDENDORF, T., and J. BUSBY. 2003. Mental time travel in animals? *Trends in Cognitive Sciences* 7 (9): 391–96.
- SUDENDORF, T., and A. WHITEN. 2001. Mental evolution and development: Evidence for secondary representation in children, great apes and other animals. *Psychological Bulletin* 127 (5): 629–50.
- . 2003. Reinterpreting the mentality of apes. In *From mating to mentality: Evaluating evolutionary psychology*, edited by K. Sterelny and J. Fitness. New York: Psychology Press.
- TATTERSALL, I. 1998. *Becoming human: evolution and human uniqueness*. New York: Harcourt Brace.
- TOMASELLO, M., B. HARE, and T. FOGLEMAN. 2001. The ontogeny of gaze following in chimpanzees, *Pan troglodytes*, and rhesus macaques, *Macaca mulatta*. *Animal Behaviour* 61 (2): 335–43.
- TOOBY, J., and L. COSMIDES. 1992. The psychological foundations of culture. In *The adapted mind: Evolutionary psychology and the generation of culture*, edited by J. Barkow, L. Cosmides, and J. Tooby. New York: Oxford University Press.
- VICK, S., and J. ANDERSON. 2003. Use of human visual attention cues by Olive Baboons (*Papio anubis*) in a competitive task. *Journal of Comparative Psychology* 117 (2): 209–16.
- WELLMAN, H. 1990. *The child's theory of mind*. Cambridge, Mass.: Bradford Books, MIT Press.
- WELLMAN, H., D. CROSS, and J. WATSON. 2001. Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development* 72 (3): 655–84.

- WELLMAN, H., and K. H. LAGATTURA. 2000. Developing understandings of mind. In *Understanding other minds: Perspectives from developmental cognitive neuroscience*, edited by S. Baron-Cohen and D. Cohen. Oxford: Oxford University Press.
- WELLMAN, H., and D. LIU. 2004. Scaling Theory of Mind tasks. *Child Development* 75 (2): 523–41.
- WELLMAN, H., and J. WOOLEY. 1990. From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition* 35 (3): 245–75.
- WILDMAN, D. E., M. UDDIN, G. LIU, L. I. GROSSMAN, and M. GOODMAN. 2003. Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: Enlarging genus *Homo*. *Proceedings of the National Academy of Sciences USA* 100 (17): 7181–88.
- WILLIAMS, G. C. 1966. *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton, NJ: Princeton University Press.
- WOODWARD, A. 1999. Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development* 22 (2): 145–60.
- YOUNG, A. W., J. P. AGGLETON, D. J. HELLAWELL, M. JOHNSON, P. BROOKS, and J. R. HANLEY. 1995. Face processing impairments after amygdalotomy. *Brain* 118: 15–24.
- ZAHN-WEXLER, C., and M. RADKE-YARROW. 1982. The development of altruism: Alternative research strategies. In *The development of prosocial behavior*, edited by N. Eisenberg. New York: Academic Press.
- ZAHN-WEXLER, C., M. RADKE-YARROW, and R. A. KING. 1979. Child-rearing and children's prosocial interactions toward victims of distress. *Child Development* 50: 319–30.

Valerie Stone, School of Psychology, Melbourne Building, University of Queensland, St. Lucia, Queensland 4072, Australia. E-mail: val.stone@psy.uq.edu.au

DO PSYCHOPATHS REALLY THREATEN MORAL RATIONALISM?

Jeanette Kennett

It is often claimed that the existence of psychopaths undermines moral rationalism. I examine a recent empirically based argument for this claim and conclude that rationalist accounts of moral judgement and moral reasoning are perfectly compatible with the evidence cited.

Introduction

In recent years there has been a surge of interest from moral philosophers in research in the social and cognitive sciences on moral development and moral reasoning. The hope has been that the emerging data can contribute to and perhaps even resolve longstanding debates in meta-ethics and moral psychology. Indeed, in a recent paper Shaun Nichols (2002) argues explicitly that the evidence from psychopathy fatally undermines the central claims made by moral rationalists.

Though I think that the empirical data warrants careful consideration by philosophers, I am concerned that philosophers (and others) might tend to extrapolate beyond the data and draw philosophical conclusions that the data do not support. First, the data which is cited may not constitute all of the relevant data and philosophers may not be best placed to assess its quality. They thus risk reaching naïve conclusions about the weight of the evidence and may too readily recruit inadequate data to support favoured positions. Second, particular care needs to be taken in applying empirical findings to philosophical accounts of moral reasoning and judgement, since these accounts are not merely descriptive. They also provide normative constraints on what is to count as sound moral reasoning and conceptual constraints on what is to count as a moral judgement. Debates on these matters may be largely unaffected by empirical evidence.

Keeping in mind these potential difficulties for setting meta-ethical debate on an empirical footing, I'll examine in some detail Nichols' case against moral rationalism. In the course of doing so, I will review empirical data not considered by Nichols. I do not argue that this data speaks decisively in favour of rationalism's claims about the nature and status of moral judgement. Rather, I conclude that it shows that the capacities which rationalists focus upon in their account of moral agency are indeed critical to human moral agency and that rationalist accounts of moral judgement and moral reasoning are empirically realisable. Rationalist accounts are perfectly compatible with the available data.

EDITOR
Anthonie Meijers, *Eindhoven University, The Netherlands*

ASSOCIATE EDITORS

Joel Anderson, *Utrecht University, The Netherlands*
Jan Bransen, *Radboud University Nijmegen, The Netherlands*
Stefaan E. Cuypers, *Catholic University of Leuven, Belgium*
Cynthia Macdonald, *Queens University Belfast, UK*
Jesse Prinz, *University of North Carolina, USA*
Beate Rössler, *University of Amsterdam, The Netherlands*
Marc Stors, *Radboud University Nijmegen, The Netherlands*
Bruno Verbeek, *Leiden University, The Netherlands*
Marcus Willaschek, *Johann Wolfgang Goethe-University, Germany*

EDITORIAL ASSISTANT

Meissa van Amerongen, *Eindhoven University, The Netherlands*

SPECIAL ADVISORS

Lynne Rudder Baker, *University of Massachusetts at Amherst, USA*
Philip Pettit, *Princeton University, USA*

BOARD OF ADVISORS

John Broome
University of Oxford, UK
Jonathan Dancy
University of Reading, UK
Martin Davies
Australian National University, Australia
Pascal Engel
University of Paris IV Sorbonne, France
Harry Frankfurt
Princeton University, USA
Jürgen Habermas
Goethe University at Frankfurt, Germany
John Haldane
University of St. Andrews, UK
Axel Homneth
The Institute for Social Research at Frankfurt, Germany
Wayne Hudson
Griffith University, Australia
Hans Joas
Max Weber Kolleg at Erfurt University, Germany
Jaegwon Kim
Brown University, USA
Adam Morton
University of Alberta, Canada
Robert Pippin
University of Chicago, USA
Marya Schechtman
University of Illinois at Chicago, USA
Michael Smith
Princeton University, USA
Frederick Stoutland
Uppsala University, Sweden
R. Jay Wallace
Berkeley University of California, USA
Susan Wolf
University of North Carolina at Chapel Hill, USA

Philosophical Explorations aims to publish outstanding articles in the philosophy of mind and action, with an emphasis on issues concerning the interrelations between cognition and agency.

Contributions are welcomed in the area of the philosophy of mind and action and related disciplines such as moral psychology, ethics, philosophical anthropology, social philosophy, political philosophy and philosophy of the social sciences. The journal also welcomes contributions of an interdisciplinary kind, establishing bridges between philosophy and, for example, evolutionary biology, neuroscience, psychology, and political science.

ISSN 1386-9795

© 2006 Taylor & Francis

PHILOSOPHICAL EXPLORATIONS

An International Journal
for the Philosophy of Mind and Action

Volume 9 Number 1 March 2006

EMPIRICAL RESEARCH AND THE NATURE OF MORAL JUDGMENT
A special issue guest-edited by Philip Gerrans and Jeanette Kennett

Editorial

Anthonie Meijers

Introduction: Is cognitive penetrability the mark of the moral?

Philip Gerrans and Jeanette Kennett

Neuro-cognitive systems involved in morality

James Blair, A. A. Marsh, E. Finger, K. S. Blair and J. Luo

The emotional basis of moral judgments

Jesse Prinz

Metaethics and emotions research: A response to Prinz

Karen Jones

The moral dimensions of human social intelligence

Valerie Stone

Do psychopaths really threaten moral rationalism?

Jeanette Kennett

Is the emotional dog wagging its rational tail, or chasing it?

Cordelia Fine

The wisdom of the pack

Neil Levy

Adaptive domains of deontic reasoning

Laurence Fiddick

As you were?

Garrett Cullity

Metaethics and the empirical sciences

Richard Joyce

1

3

13

29

45

55

69

83

99

105

117

133