

# THE ROUTLEDGE HANDBOOK OF PHILOSOPHY OF THE SOCIAL MIND

*Edited by Julian Kiverstein*

6

## PERSONHOOD AND HUMANHOOD

An evolutionary scenario

*John Barresi*

### Introduction

In recent years there has been an enormous interest in the possibility of establishing a naturalistic foundation for human morality that is based, at least in part, on an account of hominin evolution (Binmore, 2005; Boehm, 2012; de Waal, 2006; Greene, 2013; Haidt, 2012; Joyce, 2006; Kitcher, 2011; Krebs, 2011; Nichols, 2004). Each of these accounts assumes that some sort of transformation occurred from the kind of emotion-based pro-social motives that sometimes determine chimpanzee behavior in their social relations with each other to explicitly moral motivations in humans that are often guided by ethical rules and moral norms constituted by the group. For instance, Philip Kitcher contrasts psychological altruism based on sympathetic responses found in chimpanzees to the kind of normative ethical rules and principles that guide human behavior. He sees the origin of what he calls “the ethical project” in the formation of collective normative rules by early modern humans of 50,000 years ago and claims that for these small bands of hunter-gathers:

Equality, even a commitment to egalitarianism, was important. . . . In formulating the code, the voices of all adult members of the band needed to be heard: they participated on equal terms. Moreover, no proposal for regulating conduct could be accepted unless all those in the group were satisfied with it.

(Kitcher, 2011, 96)

Whether these early modern humans were as reflective and egalitarian in formulating norms as appears in this account or not, there is plenty of evidence from ethnographic reports of more recent hunter-gatherers that they are quite egalitarian, and talk a lot about, as well as jointly regulate, each other’s ethical behavior through shared norms. What is certain is that nothing like this regulation of moral behavior by group-determined ethical norms occurs in our nearest great ape relative, the chimpanzee, or in any other animal. So the puzzle of hominin evolution of morality is very much tied to the evolution of normative practice, which involves group- or culturally-based rules, whether explicitly delineated or not, that determine how one *should or must* behave with respect to others, and not merely describe how one does behave.

My goal in this chapter is to argue that one crucial difference between chimpanzees and humans is that humans conceive of themselves and others as persons and selves, and that without these concepts the normative basis of human moral life would not be possible. An essential requirement to conceive of a moral norm as applying equally to different individuals in comparable situations is not only to recognize the normative demands of the situation, but also to conceive both self and others equally as members of a class of agents whose duty it is to accede to those demands. To do this requires a concept of agent that bridges the gap between self and other. That concept for human beings is the concept of person. It is the bridge concept that makes possible normative guidance that applies to individuals as a function of roles and situations in which they find themselves, not as a function of their personal emotional preferences.

For instance, in order for the universal moral norm that one should not cause pain to another human to be experienced as a duty independent of our emotional attitudes toward particular others requires that we recognize self and all other humans equally as agents who could cause pain and recipients who could receive pain and that pain should be avoided regardless of whose pain it is. Nagel (1970) nicely captures this notion of person when he describes the situation of one individual (imagine yourself) standing on the foot of another, whether this other is a friend or complete stranger. "Recognition of the other person's reality, and the possibility of putting yourself in his place, is essential. You see the present situation as a specimen of a more general scheme, in which the characters can be exchanged" (82). In imagining the exchange, you would expect the other person to release his or her foot, not because it would reduce 'your' pain, but because it would reduce 'someone's' pain, some 'person's' pain. This is a unique aspect of the human moral order not found in other animals. Humans can conceive of themselves as just another person and that all persons should be treated equally with respect to moral norms.

This uniquely moral response is not to be confused with the kind of sympathy that is found in other animals, which is an emotional response to the expressions of others that typically applies only to kin and close associates. In such cases there is no need to imagine reversed roles. One only needs to conceive of the other's expressed pain as an extension of one's own pain or as an object of personal concern. Instead, this response in humans is based on a conception of self and other as persons, impersonal objects of moral concern. While aversion to the perceived pain of the other individual may play a motivational role here, it cannot be a form of aversion that is restricted to individuals with whom one has a natural sympathy, but one that applies uniformly to all individuals that one can conceive of as persons. The other must be experienced as another self, and their pain must be understood as comparable to the pain one would experience in their position. Without a rich capacity for perspective taking that makes possible full imagination of the reversal of positions, mere aversion to pain in another would vary with one's personal relationship to the other, and this would make it a self-interested motive, not a motive based on a conception of self and other equally as persons whose pain ought to be avoided.

This "agent-neutral" (Parfit, 1984; Nagel, 1986) and impersonal way of thinking about and experiencing moral motivations requires a concept of person that contrasts with an "agent-relative" and personal way of experiencing motivations typical in other animals. In his attempt to justify the "possibility of altruism", Nagel (1970) made a formal distinction between motives or reasons that are personal and apply to a particular person and his or her relationships with others, and impersonal reasons that apply to persons in general vis-à-vis other persons. Based on Nagel's distinction, Parfit (1984) identified ethical theories as agent-relative (e.g. egoism) and agent-neutral (e.g. utilitarianism). Nagel (1986) later adopted Parfit's terms and wrote:

If a reason can be given a general form which does not include an essential reference to the person who has it, it is an *agent-neutral* reason. . . . If on the other hand the

general form of a reason does include an essential reference to the person who has it then it is an *agent-relative* reason.

(152–153)

While Nagel (1970, 1986) hoped to distinguish between personal subjective motives from objective reasons with this distinction, this is not my motivation for adopting this terminology in the present chapter. My interest is in using the terminology to distinguish a naturalistic division between the kind of agent-relative motives that apply to most animals, including chimpanzees, and the kind of agent-neutral motives that apply to humans. While in many circumstances human motives do not differ in kind from motives that we find in chimpanzees and other animals, at least in normative circumstances, human motives are governed by our conception of self and others as persons. And my main concern here is on the evolution of this way of conceiving self and other and on its role in human social life.

In what follows I will argue first that it was the adaptive need for a high level of cooperation that caused early hominins to acquire the concepts of person and self in thinking about cooperative activity. Acquiring and understanding the relationship between these two concepts depended on an ability to conceive of the point-of-views of others in the same representational form as one conceived of one's own point-of-view. The outcome was that they were able to think about the intentional activities and interests of self and other in a single common format that applied uniformly to self and others and could engage in normative guidance based on agent-neutral situational rules generated within one's group, not just on personal relations. I will then provide a 'how possibly' story of this evolution with a particular focus on the role that reciprocal altruism played and how our sense of justice emerged in this process. Second, I will argue that chimpanzees and probably other organisms do not conceive of themselves and others equivalently as persons and selves and that this results in a distinction between us and other organisms in our capacity for agent-neutral thinking. Third, I will show how humans are distinguished from other animals by their early developmental conception of self and others as persons and selves. These concepts become available in the second year of life and go through several important stages in development that are crucial to our way of life as cooperative organisms.

### **The evolution of cooperation in hominins and the adaptive function of our concept of person**

I propose that it was the adaptive need for increasing levels of cooperation at a group level that caused early hominins to acquire and develop the concepts of person and self in thinking about cooperative activity. While extending altruism through the capacity to follow orders as suggested in Kitcher's (2011) account of normative guidance is part of the story, it leaves out both motivational and conceptual resources necessary to create and follow the cooperative norms upon which those orders are based. I believe that the critical move in hominin evolution was engaging in a form of cooperation that required shared intentions that governed social behavior in a more general manner, one where each individual in a group had to take into account the view of the group as a whole when evaluating their own and other individuals' actions. The outcome was the formation of a concept of person to apply to self and others within the social group, and a notion of what persons in various roles were expected to do. Rather than merely extending an agent-relative perspective on altruism based on natural sympathy with a capacity to follow commands, normative guidance involved shifting to an agent-neutral way of thinking about of the intentional perspectives of self and other and a greater focus on the role that group-based cooperative goals played both in generating and following rules.

What drove this adaptive need for a rich form of cooperation in early hominins and how was it solved? As to the source, there appear to be two main answers in the literature: (1) the need to engage in cooperative foraging, in particular group hunting and food sharing of large animal kills; and (2) the need to engage in cooperative breeding, which included longer periods of dependent infancy and childhood to develop skills needed for survival in varying physical and increasingly complex cultural environments (Chapais, 2008; Hrdy, 2009; Sterelny, 2012). With respect to how these needs were satisfied, two evolutionary mechanisms were particularly important in making this intense form of cooperation possible: kin selection and reciprocal altruism (e.g., Binmore, 2005; Sterelny, 2012). How it all started is uncertain, but when hominids entered the savannas during the late Miocene there was pressure to live in cooperative groups for the purposes of joint defense against predators and joint foraging of dispersed sources of food. Early developed bipedalism made distance traveling possible for the evolving hominins and created the opportunity to use hands for a variety of other purposes, including for the creation and modification of tools used for defense and foraging, and for communication of cooperative intentions.

The evolutionary mechanisms of kin selection and reciprocal altruism were important for ramping up high levels of cooperation to survive and evolve in these novel circumstances. Exactly what happened and when it happened is unclear but, with respect to kin selection, there is reason to believe that hominin cooperation among kin shifted from maternal-only relations, as currently found in our closest ape relatives, to maternal and paternal relations, through bonding between particular males and females that made possible recognition by males of their offspring, thus warranting greater investment in their care. This new arrangement increased the role of bilateral kin relations, both close and distant, which opened the door to longer periods of child development and to the exchange of marriage partners between kin groups (Chapais, 2008).

With respect to reciprocal altruism, cooperative foraging required tools, coordinated activity, advanced planning, reliable partners, and group-level food sharing of large animals. Group-level cooperative hunting not only involved kin but also non-kin, so free riding by individuals who sought rewards without paying the costs of cooperation became a serious problem requiring normative control. In order to control free riding social contract dynamics eventually led to group structures with a flat dominance hierarchy (egalitarianism), normative rules for sharing, and effective forms of group-based punishment (Binmore, 2005; Boehm, 2012; Boyd & Richerson, 1992; Skyrms, 2004; Trivers, 1971). Coordinated hunting that required evaluation of potential partners (both kin and non-kin), shared intentions, and multiple roles, as well as future-planning for and commitment to shared distribution of uncertain but high-density food sources, required a theory of mind and a temporally extended sense of self and other. Communication and sharing of intentions and knowledge (the basis of language) with non-kin as well as kin required generalized forms of reciprocation. Exchange of marriage partners, within and between bands involving distant kin and non-kin, eventually created long-term reciprocal bonds between non-kin groups. Thus small bands eventually became integrated into tribes that would compete with each other for resources, stimulating an intergroup dynamic for both group selection of cooperative genes and cultural selection of variations of group and individual behavior.

The capacity for agent-neutral ways of thinking about the intentional activities of self and other became particularly important in activities involving reciprocation, where balancing of costs and benefits for each of the individuals involved required tracking inequities that might affect fitness, and a metric for uniform calculation of costs and benefits to different individuals. While kin selection and inclusive fitness can function well with agent-relative motivations to

act altruistically toward relatives – though with variation due to degree of relationship – altruism to non-relatives requires close attention to costs and benefits to different individuals that can be calculated in an agent-neutral way. So the development of increasingly abstract agent-neutral concepts representing the intentional activities of individuals within the group must have been especially driven by cooperative activities involving non-relatives, though even with relatives, attention to relative costs and benefits in an agent-neutral manner was required when costly forms of altruism were involved.

To better appreciate the need for abstract forms of agent-neutral thinking, consider a likely scenario for the advanced hominins (perhaps as early as *Homo ergaster* or *Homo erectus*) who engaged in hunting for large animals. The group of hunters had to decide on when and what to hunt, which weapons collectively created and owned to bring with them, and how to divide their group into smaller divisions in order to cover the area where they might find the best target animals. This advanced planning, which occurred at a collective level, required temporally extended notions of self and other and the use of future-oriented imagination based in part on past memories of successful hunts, an understanding of the skills of each of the participants, and some insurance or trust that each of the participants would put in roughly equal effort in hunting. Moreover, when some animal was killed, the participants had to trust that all members of the hunting party would get equal access to the best parts of the animal on site to eat without significant dispute, and then work together to bring the animal back to the home base to share, not only with relatives of the hunters, but with the whole group in a sufficiently egalitarian way that the fitness of all individuals in the group would be enhanced to facilitate group survival.

The calculation of costs and benefits in this one scenario (and there are many others that could be described) in a way that would maintain any degree of equity would probably be impossible if done in an agent-relative way, requiring each individual to make calculations vis-à-vis every other individual in the group with respect to the series of events in this scenario and to negotiate an agreement to follow a common plan. Agent-relative thinking and motivation would undoubtedly lead to a breakdown in cooperation at multiple points in the scenario. It is unlikely that cooperative habits and collective thinking that would be required to engage in this scenario could ever evolve with agent-relative thinking, as the limitations of chimpanzee group hunting illustrates (Tomasello & Vaish, 2013; but see Kaufmann, this volume, for an opposing view of chimpanzee cooperation in hunting). Even with agent-neutral thinking, calculating costs and benefits to individuals would be difficult if each hunt was treated as a particular event. But agent-neutral thinking based on roles, temporally extended understanding of self and other including personality traits, and life-course social identities that can be abstractly conceived, along with norms that govern behavior as a function of these categories, afford the possibility of gradual acquisition of norms for scenarios of this sort as integrated within an ever evolving group cultural context.

Recently, Ken Binmore (2005) has proposed a social contract version of game theory that captures some of what would be required here. Although his illustrations deal only with indefinitely repeated 2-person games of various sorts, the basic idea helps to understand what agent-neutral thinking can do to facilitate the gradual development of equilibria with increased collective payoffs. In his model of social justice, he assumes that each individual not only has a personal utility function, but also creates an accurate utility function for their partner for interpersonal comparisons in various games with multiple equilibria. Some of these equilibria are more optimal than others for them collectively. He proposes that the capacity to create interpersonal utility functions is based on empathy that derives originally from kin selection but is then generalized to non-kin. However, this is an unlikely source, as kin selection only requires

agent-relative thinking that will not produce uniform representations of costs and benefits for self and other. Instead, an agent-neutral form of thinking is required, one that would have had its basis in reciprocal altruism. The important suggestion that he makes is that the capacity to search for and adopt better equilibria when bargaining can be represented metaphorically using Rawls' (1971) notion of the original position, where participants bargain as if they did not know which person they would turn out to be with respect to the game at hand and use the interpersonal utility functions to make optimal decisions. Binmore proposes that natural justice can evolve within this framework from one equilibrium to another through time, where each equilibrium sets the norms for the participants at one time, which remain stable until some destabilizing change occurs.

Without getting into more details of this theory, what appears clear is that it invokes a form of agent-neutral thinking and understanding of self and other equally as persons in various roles in a stable social network with varying power relations. Equity and equality don't just happen here, but because of our capacity for thinking of our selves and others equally as persons in agent-neutral ways, progress is possible toward better cooperative equilibria approaching egalitarian ideals. Use of the original position in Binmore's model is similar in some respects to Nagel's notion of person as a basis for rational reflection about moral judgments. In both models there is an ideal component that purports to be the basis of rational decision-making. However, Binmore's is more naturalistic, as he claims that rational self-interested actors would progress to better equilibria under his assumptions. However, hidden here is the assumption that rational agents will act in an agent-neutral manner, not in an agent-relative way. Without the capacity to think in an agent-neutral manner about costs and benefits for self and other, the idealization of the original position would not work. Instead, agent-relative motivations would prevent successfully finding cooperative equilibria and result instead in more competitive dominating outcomes of the sort found in chimpanzees, and high levels of cooperation would not occur. Our own high levels rely on agent-neutral thinking and the norms generated over time to legitimize and coordinate our cooperative behavior. While we don't always behave according to culturally constituted agent-neutral norms, it remains the main means by which our society is organized and holds us together in complex collective agreements. And if agent-neutral ways of thinking are required for cooperation of this sort, then I argue that it was evolution of our concept of self and other as persons that was at the root of the social contracts in which we currently engage as well as for the ideals of social justice that we hope to achieve in the future.

### The phylogenesis of persons and selves

In common language and experience we use the English terms 'self' and 'other' to describe the same kind of thing, an individual human being or person. Like 'I' and 'you', they are deictic terms that shift with the user. In order to use these terms, we must be able to recognize that you and I are both persons – that we are equivalent in this way. But we experience the personhood of self and other in different ways. Our experience of our own personhood is a first-person experience of our self, while our experience of each other's personhood is from a third-person perspective. These two forms of experience are intrinsically different but can be conceptually connected. Our first-person experience of our own psychophysical activity, what Barresi and Moore (1996) call intentional relations, is focused outward on the object of our activity, whether it be a goal directed action, an emotional relation to another individual or object, or an epistemic relation to an object or situation. Our experience of 'self' is only in our embodied relations to these activities and objects; it is not typically an object of our attention,

but implicit in our relational experience of other objects (see Musholt, 2015). By contrast, our third-person experience of another individual's intentional relations focuses on the individual as an animate agent, and less so on the object of these relations. We see the other individuals' physical movements, emotional expressions, and direction of gaze; and often we must infer the objects of these activities by attending to the locations toward which these actions are directed. Yet, despite these different perspectives and information available about the activities of self and other, as adult humans we have no difficulty recognizing both self and other as agents of the same kind – as selves and persons – and can ascribe to our activities the same intentional concepts. In the next section we will view human development of these concepts and how they relate to our notions of self and persons. However, here we focus on other animals, and, in particular, those in the hominid line.

In considering personhood from a phylogenetic perspective, the key issues are: When does a common conception of self and other as individual embodied agents engaged in intentional relations that are understood in the same way first appear? Why does it first appear at that time? And is this common conception as elaborate as our own human concepts of persons and selves?

Barresi and Moore (1996) have argued that most animals represent their own intentional relations in a first-person format and the intentional relations of others in a third-person format, and as a result self and other are represented quite differently. However, certain highly social animals may be capable of representing some activities of self and other in a common format. One form of evidence indicating this possibility occurs when an animal is able to recognize itself in a mirror. While this test does not guarantee an equivalent concept of itself and others, it does indicate that the animal can treat the reflections in a mirror of itself and another equally as indicating the current physical appearance and activities of a particular individual. They appear here to adopt a third-person perspective on self through the use of the mirror.

Animals that pass stringent forms of the mirror test include great apes (gorillas, orangutans, chimpanzees, and bonobos), cetaceans (e.g., dolphins and killer whales), elephants, and at least one species of bird (magpie) (Gallup, Anderson, & Platek, 2010). All of these species have relatively large brain/body ratios. Moreover, there is evidence that most of them have the converse capacity to imagine the first-person perspective of others, something that again is unusual among non-human animals. For instance, great apes, cetaceans, and elephants show fairly strong evidence of empathy, not only responding to the expressed distress of kin, but also to unexpressed situational needs even of non-kin (de Waal, 2008). There is also evidence to suggest that great apes and dolphins can imagine the visual viewpoint and knowledge of others. Taken together, these findings suggest that such animals can, in some presumably advantageous social circumstances, conceive of both self and other from a first- and a third-person point of view. These capacities can be useful, for instance, in acquiring a novel skill through imitation, or in taking advantage of the false knowledge of the other in competition for a hidden resource.

An important question that remains is whether these two views of self and other are integrated in these organisms. If not integrated, then there may not be a single form of representation, but two separate more limited forms that can be applied to both self and other. Particularly relevant for evaluating whether the chimpanzee's concept of self and other are similar to our notion of person are the circumstances in which chimpanzees engage in joint activities. Although chimpanzees live in social groups and sometimes work cooperatively for a common goal (e.g., territorial defense, hunting for monkeys), for the most part they pursue goals independently of each other, and often compete with each other for desirable objects. Recent research has tried to investigate chimpanzee cooperation and to compare it to that found in young children (Fletcher, Warneken, & Tomasello, 2012; Rakoczy et al., 2014; Tomasello & Vaish, 2013). At least in experiments scaffolded by humans, chimpanzees can learn to

cooperate with a partner in achieving joint rewards. But in doing these tasks they seem to learn only about their own role, and do not acquire any useful knowledge of the role of their partner. Thus, when the chimpanzees switch roles with their partner, this does not increase their speed in learning the task. By contrast, young children do increase speed upon switching roles in these experiments, which suggests that they represent not only their own task, but also that of their partner in a common framework. These results suggest that the chimpanzee is representing the joint activity only from a first-person point of view of their own task and a third-person point of view of the task of the other individual. This indicates that the chimpanzee has only an agent-relative perspective of the tasks of self and other while the child represents the tasks of both agents from an agent-neutral intentional perspective.

Other experiments involving cooperation and altruism show that chimpanzees in cooperative tasks only pursue outcomes that are in their own interest and show no interest whatever in whether their partner will be rewarded or not. When they have the option of choosing one response that will reward both, or another that will only reward themselves, they are equally likely to choose one or the other response (Jensen et al., 2006), whereas preschool children prefer to reward both self and another. Indeed, children are willing to forgo a current reward for self in order that self and other both gain rewards in the future (Thompson, Barresi, & Moore, 1997). Moreover, in tasks that require cooperation, chimpanzees never attempt to communicate with each other to encourage their partner to do the necessary complementary actions. It's as if they do not represent the task as a joint activity at all, but see it only from their own relative point of view, seeing what they have to do, and observing the actions of others as if the other is pursuing its own ends. Since the idea of a common goal governing their joint activity is apparently absent in their thinking of the task (Tomasello, 2014), there is reason to think that whatever concepts that chimpanzees use to compare self and other are not of the same kind as our human concept of person.

The lack of interpersonal relations of a cooperative nature in joint actions may limit the chimpanzee's ability to represent intentional relations of self and other using a single uniform concept that integrates first- and third-person perspectives. If shared intentional relations ground human understanding of our selves and others as persons, then the limited bridge between self and other that we see in chimpanzee behavior may not be enough. Without integration of perspectives based on joint activity that begins in human infancy, chimpanzees cannot acquire an agent-neutral concept of intentional agent that they can apply equally to self and other in cooperative activities, thus limiting them to an agent-relative view of those activities. Other species, in particular some cetaceans, are more cooperative than chimpanzees and on several measures seem to show a more advanced understanding of self and other as equivalent intentional agents. But comparing humans to cetaceans would do little to answer the question of the origin of the concept of person in the hominid line.

### **Personhood, self-reflection, and the development of agent-neutral perspectives**

Given the differences in experience of intentional relations from first- and third-person points of view, how is it that human beings come to understand both self and others as agents of the same kind and are able to ascribe at least some psychological attributes equally to self and others? The two major theories of how we understand mental phenomena in ourselves and others – simulation theory (ST) and the theory theory of mind (TT) – have difficulty explaining the ease with which children acquire this understanding. ST gives precedence to first-person information and representations of mental states and generalizes from self to others.



By contrast, TT gives precedence to third-person behavioral information and then generalizes theoretical concepts of mind based on this information to self. Both of these approaches suppose a dualist conception of the relationship between mental phenomena and their behavioral expression, making it difficult to bridge the gap between them based on radically different sources of information about their relationship from self and other.

An alternative approach is to focus on the person rather than on mental states (Barresi, Moore, & Martin, 2013; Dow, 2012; Newen & Schlicht, 2009). The inspiration for these theories is Peter Strawson's (1959) non-dualist account of persons. Strawson insists that we cannot do without two ways of viewing ourselves and others, a first- and a third-person perspective, and that whatever psychological concepts we use to describe self must also be equally and *unambiguously* useful in describing others, though based on *quite different criteria*. He views the concept of person as a primitive and essential one that necessarily precedes any notion of a conscious or mental self. He points to a number of contradictions that arise when we attempt to view ourselves as conscious selves, or minds, on criteria that are independent of our bodies and also try to attribute analogous conscious selves to others based on their behavior. These lead to various forms of dualism such as those that appear in TT and ST.

Strawson made several suggestions on how we might acquire psychological concepts, despite the different perspectives that we have of our own and another person's activities. In particular, he noted that some activities, like walking, involve behavioral and mental aspects so intermingled that these activities can be readily bridged in understanding from a first- and third-person perspective. He also noted that some activities, such as group sports, are joint activities with common goals, and he suggested that a notion somewhat akin to that of a 'group mind' could play a role in understanding these activities where 'we' rather than 'you' or 'I' have a singular goal to be achieved. In this case, there is no issue about whether there are different concepts being applied to self and other based on different criteria. It is a single concept of what *we* are doing that is applied, but understood in a first-person manner for self and in a third-person manner for others. Thus, there is a perfect match in the intention as experienced in the first-person and as experienced at the very same time in the third-person, and hence no gap in the content of the intentional state attributed to self and other based on different criteria.

Barresi et al. (2013) have shown how these ideas of Strawson are congruent with events that occur in early child development. Of particular importance for the present chapter, which focuses on the evolution of cooperation, is the second suggestion. One of the remarkable features of early human development compared to that of other animals is that human infants engage with adults, and later on with other children, in joint intentional activities directed at common goals, which often involve mutual imitation of actions, joint attention, and by-directional communication about objects and goals. As pointed out in the previous section, there is little evidence of cooperative activity of this sort in other existing primates (Tomasello, 2014; Tomasello & Vaish 2013). Because this activity involves shared intentional relations between the infant and the adult, the infant can experience the common mental and physical aspects of the activity from both a first- and a third-person perspective and can link them as involving the same activity. Thus it can represent the activity of self and other in a common format, one that applies to their joint activity, despite the different types of information presented for self and other. This format associates the first-person perspective of self with the third-person perspective of the other into a concept of their joint activity where both perspectives are included. Moreover, this cooperative activity often involves complementary roles that are reversed over time, which also facilitates relating the two sources of information about the same activity and bringing them into a common format that can eventually be applied to each of the individuals separately and not just in joint activity.

AuQ13

As a result of this early joint intentional activity the child comes to understand a variety of intentional relations in a single format that can be applied to self and other, not only when they are engaged in a common activity, but also when they are engaged in different activities. By the end of the second year toddlers have come to recognize themselves and others as distinct agents, who can pursue *different* activities. Yet the toddler can conceive of these different activities of both self and other as activities of an embodied agent, a person and self. When the toddler observes the actions of another individual she can imagine the first-person perspective of those actions, as if she shared in this activity with the other individual. Thus, prior experience of joint intentional activities provides a basis for recognizing the meaning of these actions from the first-person perspective of the other individual. Conversely, when the child engages in her own actions, she can imagine the third-person perspective of another individual observing her, and joining with her in the actions, and can understand, in a reflective manner, her own activity as an embodied agent or person no different from any other person. Thus, the first- and third-person aspects of intentional activity can now be unified in interpreting actions of individuals, as well as in conditions of joint action. Both self and other are now perceived from a meta-perspective that represents them both as embodied agents that are persons and selves, each of which can be viewed from an 'objective' or 'allocentric' third-person perspective and each attributed a 'subjective' or 'egocentric' first-person point of view.

In accordance with Strawsonian requirements, there is a perfect symmetry in the representation of intentional actions ascribed to both self and other. The child now understands the other as another self, and the self as another other. As a consequence the child is now able to use deictic terms like 'I' and 'you' in an appropriate manner, and begins to experience forms of self-consciousness like embarrassment that she could not previously experience because they require a level of representation of the self as a whole agent who is the possible object of another's attitudes.

When engaged at this age in joint activity, the toddler can now readily shift roles, because she can imagine the first-person perspective of the other individual in complementary positions in any activity. Instead of experiencing one's own role only in a first-person format, and the role of the other only in a third-person format, the child now represents both roles in an integrated format with both first- and third-person aspects. Thus, the child sees joint activity as that of two agents engaged in shared intentional activities with common goals, where both agents and their independent roles are understood so as to allow the child to take on either role if that were required (Rakoczy et al., 2014).

From the ages of 2 to 4, children acquire increasing skills in thinking of themselves and others as embodied, intentional agents engaged in activities that involve complementary roles that are often governed by various conventional and moral rules or norms. They become skilled in thinking of these intentional activities in agent-neutral terms that apply norms or rules as a function of the roles that are involved in these activities, whether culturally constituted and acquired from adults or created in collaborative play with other children. They also acquire the necessary executive skills to regulate their actions with respect to these norms. Thus, these children acquire the capacity for normative guidance recognized in Kitcher's (2011) theory. Indeed, children of this age are sticklers about playing by rules and insist that others as well as self play by them. They are also able to distinguish between conventional and moral norms, the latter being those that they believe hold universally, while the former are arbitrary and restricted to smaller groups (Nichols, 2004; Rakoczy & Schmidt, 2013; Tomasello & Vaish, 2013).

Although the 2- and 3-year-old has acquired a concept of person and self and is capable of thinking of self and other in agent-neutral ways, especially when engaged in norm-governed activities, this conception of person and self is limited to the here and now, or extended only

with respect to well-known routines. There are two more major advances that are needed for children and adolescents to achieve an adult concept of person (Barresi, 1999). These two advances are the concept of a temporally extended person and self, and a life-course narrative identity. Both of these concepts are necessary in order to recognize and perform adult activities in an agent-neutral manner.

A major change occurs during the fourth and fifth years of development. It is at this time that self-reflection enters more fully into the temporal domain and the child becomes capable of moving imaginatively not only across space from person to person in present time but also across time to past and future person positions (Barresi 2001; Moore & Lemmon 2001). The child is able to conceive of its own past and future representations of reality as distinct from its present representations, and begins to appreciate itself as well as others as selves extended in time. Before this time, experiences unfold but are not connected together into an autobiographical stream. Now, retrospective memory and anticipation of the future has this structure. Correlated with these skills are the executive capacity to act for future rather than for present goals and the abilities to understand false belief and what is called level-two perspective taking that fall collectively under the concept of representational theory of mind (Moore, Barresi, & Thompson, 1998).

The acquisition of the concept of an extended self makes possible cooperative and moral activity that extends over time. Promises can be made, remembered, and kept, at least over short periods of time. However, over longer periods and different situational contexts there may be a lack of stable moral outlook. Not until adolescence is there an attempt to maintain a stable moral stance across time and through a variety of situations. Consider, for instance, some developmental research on the keeping of promises to friends. Keller (2004) has shown in a cross-cultural study a developmental trend in the likelihood that a child or adolescent will think a same-age actor will keep a promise to be with a friend in a dilemma that requires a choice between keeping the promise versus accepting an invitation to go to a movie with another child, who is new in class. Most Icelandic and Chinese 7- and 9-year-olds thought that the actor would choose to go to the movie, while most 15- and 19-year-olds thought that the actor would keep the promise to be with the friend. Twelve-year-olds were transitional. In justifying their decisions, young Icelandic children thought that going to the movie would be more fun, though they also thought that the right thing to do would be to keep the promise to be with the friend. The Chinese children justified the same action by saying that it is right to be nice to the stranger child. However, adolescent participants in both cultures agreed that it was important to keep the promise to a friend and thought that this was the right thing to do because otherwise the actor would not be perceived as a reliable and trustworthy person, and would not really be a friend.

This acquired concern over one's moral identity that persists through time and varying situations is one outcome of a process of identity formation that typically begins in adolescence (McAdams, 1990). Young adults can now conceive of themselves and others as persons with consistent individual personalities as well as life-course identities constructed out of those available within their culture, thus conceivable in agent-neutral ways. Being known as a trustworthy person becomes an essential personality attribute for long-term cooperative relationships with others such as friendships, marriage, and careers. Importantly, young adults can now formulate and regulate their behavior by abstract moral principles and norms that apply across most situations. They can also join with others in formulating moral norms as well as other agent-neutral conventions of the society in which they live, and can play their part in the general government of each other's behavior in accordance with those norms. To the extent that such norms are based on principles of equality and equity, they contribute to an egalitarian society. However, variation

in social identities brings with it asymmetries in power and influence, and non-egalitarian norms can persist because individuals in power take advantage of their social identities to maintain those asymmetries. While agent-neutral thinking of all members of the society as persons promotes egalitarian ideals, specific identities and power relationships undermine this kind of thinking with more agent-relative motivations.

Overall, my proposal is that as our concepts of person and self develop we acquire a capacity for wider agent-neutral forms of representation. Compared to the more advanced stages, earlier concepts of person and self are more limited and agent-relative. Agent-neutral cooperative and moral activity at each stage can only go so far without the wider perspective that is more inclusive. The narrowest stage does not involve a concept of person and is purely agent-relative. This is the stage that most animals are at, even chimpanzees. If it were not for the adaptive need for more intense forms of cooperation than that required for chimpanzee life our ancestors may never have made the leap to the kind of agent-neutral ways of thinking about self and other equally as persons that provides a necessary conceptual capacity to ground human moral life.

## References

- Barresi, J. (1999). On becoming a person. *Philosophical Psychology*, 12, 79–98.
- . (2001). Extending self-consciousness into the future. In C. Moore and K. Lemmon (Eds.), *The Self in Time: Developmental Perspectives* (pp. 141–161). Hillsdale, NJ: Erlbaum.
- Barresi, J. & Moore, C. (1996). Intentional relations and social understanding, *Behavioral and Brain Sciences*, 19, 107–122.
- Barresi, J., Moore, C. & Martin, R. (2013). Conceiving of self and others as persons: Evolution and development. In J. Martin and M. Bickhard (Eds.), *The Psychology of Personhood: Philosophical, Historical, Social-Developmental, and Narrative Perspectives* (pp. 127–146). Cambridge, UK: Cambridge University Press.
- Binmore, K. (2005). *Natural Justice*. Oxford: Oxford University Press.
- Boehm, C. (2012). *Moral Origins: The Evolution of Virtue, Altruism, and Shame*. New York: Basic Books.
- Boyd, R. & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13, 171–195.
- Chapais, B. (2008). *Primeval Kinship: How Pair-bonding gave Birth to Human Society*. Cambridge, MA: Harvard University Press.
- de Waal, F.B.M. (2006). *Primates and Philosophers: How Morality Evolved*. Princeton, NJ: Princeton University Press.
- . (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, 59(2008), 279–300.
- Dow, J. (2012). On the joint engagement of persons: Self-consciousness, the symmetry thesis and person perception. *Philosophical Psychology*, 25, 49–75.
- Fletcher, G., Warneken, F. & Tomasello, M. (2012). Differences in cognitive processes underlying the collaborative activities of children and chimpanzees. *Cognitive Development*, 27, 136–153.
- Gallup, G. G. Jr., Anderson, J. R. & Platek, S. M. (2010). Self-recognition. In S. Gallagher (Ed.), *The Oxford Handbook of Self* (pp. 80–110). Oxford: Oxford University Press.
- Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between us and Them*. New York: Penguin Press.
- Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Pantheon.
- Hrdy, S. (2009). *Mothers and Others*. Cambridge, MA: Harvard University Press.
- Jensen, K., Hare, B., Call, J. & Tomasello, M. (2006). What's in it for me? Self-regard precludes altruism and spite in chimpanzees, *Proceedings of the Royal Society: B*, 273, 1013–1021.
- Joyce, R. (2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.
- Keller, M. (2004). Self in relationship. In D. K. Lapsley and D. Narvaez (Eds.), *Moral Development, Self, and Identity* (pp. 267–298). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kitcher, P. (2011). *The Ethical Project*. Cambridge, MA: Harvard University Press.
- Krebs, D. L. (2011). *The Origins of Morality: An Evolutionary Account*. Oxford: Oxford University Press.

- McAdams, D. P. (1990). Unity and purpose in human lives: The emergence of identity as a life story. In A. I. Rabin, R. A. Zucker, R. A. Emmons and S. Frank (Eds.), *Studying Persons and Lives* (pp. 148–200). New York: Springer.
- Moore, C., Barresi, J. & Thompson, C. (1998). The cognitive basis of future-oriented prosocial behavior. *Social Development*, 7, 198–218.
- Moore, C. & Lemmon, K. (Eds.) (2001). *The Self in Time: Developmental Perspectives*. Hillsdale, NJ: Erlbaum.
- Musholt, K. (2015). *Thinking about Oneself: From Nonconceptual Content to the Concept of a Self*. Cambridge, MA: MIT Press.
- Nagel, T. (1970). *The Possibility of Altruism*. Oxford: Oxford University Press.
- . (1986). *The View from Nowhere*. Oxford: Oxford University Press.
- Newen, A. & Schlicht, T. (2009). Understanding other minds: A criticism of Goldman's simulation theory and an outline of the person model theory. *Grazer Philosophische Studien*, 79, 209–242.
- Nichols, S. (2004). *Sentimental Rules*. Oxford: Oxford University Press.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Rakoczy, H., Gräfenhain, M., Clüver, A., Schulze Dalhoff, A. C. & Sternkopf, A. (2014). Young children's agent-neutral representations of action roles. *Journal of Experimental Child Psychology*, 128, 201–209.
- Rakoczy, H. & Schmidt, M. (2013). The early ontogeny of social norms. *Child Development Perspectives*, 7, 17–21.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Skyrms, B. (2004). *The Stag Hunt and the Evolution of Social Structure*. New York: Cambridge University Press.
- Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. Cambridge, MA: MIT Press.
- Strawson, P. F. (1959). *Individuals*. New York: Taylor and Francis.
- Tomasello, M. (2014). *A Natural History of Human Thinking*. Cambridge, MA: Harvard University Press.
- Tomasello, M. & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology*, 64, 231–255.
- Thompson, C., Barresi, J. & Moore, C. (1997). The development of future-oriented prudence and altruism in preschool children. *Cognitive Development*, 12, 199–212.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35–57.