



7 Conceiving of self and others as persons: evolution and development

John Barresi, Chris Moore, and Raymond Martin

In common language and experience, at least in English, “self” and “other” are opposing terms for the same kind of thing, an individual human being or person.¹ Like “I” and “you,” they are deictic terms that shift with the user. I am a “self” to me and an “other” to you, and you are a “self” to you and an “other” to me. In order for us 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 our personhood is experienced differently by each of us. Your experience of your personhood is an experience of your self and my experience of my personhood is an experience of my self. Our experience of each other’s personhood is that of an-other person.

Given the differences in our experience of self and other, how is it that we come to understand both selves and others as persons and can learn to apply deictic terms such as “I” and “you” so easily that even 2-year-olds can have a conceptual understanding of themselves and others as the same kind of things, selves and persons, and are able to ascribe at least some psychological attributes equally both to themselves and others? The two major current 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, given how the asymmetry in our access to information about mental states of self and other is dealt with in these theories.² As a consequence, recently there has been a renewed interest

¹ We wish to thank Jack Martin and Mark Bickhard for their editorial assistance. Separate grants from Social Sciences and Humanities Research Council (SSHRC) of Canada to John Barresi and Chris Moore supported research for this chapter.

² D. Premack and G. Woodruff, Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1 (1978) 515–525, originated the label “theory of mind” (ToM); A. Goldman, *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading* (Oxford University Press, 2006) provides a recent review of the two major theories. A third type of theory focuses on intersubjectivity and interpersonal relations. Our own view falls within this third approach; see, J. Barresi and C. Moore, Intentional relations and social understanding. *Behavioral and Brain Sciences* 19 (1996) 107–122, and, especially, The neuroscience of social understanding. In J. Zlatev, T. Racine, C. Sinha,

in theories that focus on the person rather than on mental states. These theories maintain that there is a constitutive link between our ability to apply psychological terms to self and other. We could not learn to apply mental concepts to self without also learning how to apply them to others. Likewise for the reverse relation, our ability to recognize psychological states of others is intimately tied to our ability to recognize them in ourselves. These person theories maintain that there is a direct non-causal, necessary connection between mental events and their expression. Thus, mental events are not hidden within the organism with only contingent causal relations to their behavioral expressions. Rather, they are transparent in their expression. While the enormous figure of Wittgenstein stands in the background of this alternative approach, much of the inspiration for these theories is Peter Strawson's non-dualist account of persons.³

Strawson insists that we cannot do without two ways of viewing ourselves and others, a first- and a third-person perspective, and that whatever concepts we use to describe self must also be equally useful in describing others. In his view, this requires that we understand our selves and others as persons, where persons are objects with material properties that apply to all objects (M-predicates) as well as psychological properties that apply specifically to persons (P-predicates). 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. Such attempts always give rise to some form of a mistaken dualism in describing persons, though not always dualisms of substance. Unless material and mental predicates can both be ascribed to the same individual entity we can make no coherent sense of their application equally to self and other.

Strawson provides a useful summary of the problems that arise when one tries to provide a general theory of mind that gives priority to the first-person perspective of self (like ST) and is generalized to others, or

and E. Itkonen (eds.) *The Shared Mind: Perspectives on Intersubjectivity* (Amsterdam: John Benjamins, 2008) 39–66.

³ See A. Avramides, *Other Minds* (London: Routledge, 2001); J. L. Bermúdez, *The Paradox of Self-consciousness* (Cambridge, MA: MIT Press, 1998); J. Dow, On the joint engagement of persons: Self-consciousness, the symmetry thesis and person perception. *Philosophical Psychology* (in press); A. Newen and T. Schlicht, Understanding other minds: A criticism of Goldman's simulation theory and an outline of the person model theory. *Grazer Philosophische Studien* 79 (2009) 209–242; Seemann, A. Person perception. *Philosophical Explorations* 11 (2008) 245–262; and P. F. Strawson, *Individuals* (New York: Taylor and Francis, 1959).

gives priority to the third-person perspective of others (like TT) and is generalized to self:

Just as there is not in general one primary process of learning, or teaching oneself, an inner meaning for predicates of this class (P predicates), then another process of learning to apply such predicates to others on the strength of correlation, noted in one's own case with certain forms of behavior, so – and equally – there is not in general one primary process of learning to apply such predicates to others on the strength of behavior criteria, and then another secondary technique of exhibiting a new form of behavior, viz. first-person P-utterances.⁴

Strawson says that both of these pictures are refusals to note the unique logical character of P-predicates – that these predicates, or properties of persons, are known and ascribable in different ways from the first- and third-person perspective:

It is essential to the character of these predicates that they have first- and third-person ascriptive uses, that they are both self-ascriptible otherwise than on the basis of observation of behavior of the subject of them, and other-ascriptible on the basis of behavioral criteria . . . In order to understand this type of concept, one must acknowledge that there is a kind of predicate which is unambiguously and adequately ascribable both on the observation of the subject of the predicate and not on this basis in one's own case.⁵

Strawson says that without the possibility of concepts of this type we could have no concept of person that applies equally to self and other.

Like Wittgenstein before him, Strawson claims that there are non-causal internal relations between the first-person perspective and information that we have about P-properties in our own case, and the third-person perspective and information that we have about P-properties in the case of others, so that we learn to apply these concepts by linking experiential information associated with both points of view. However, not all instances of P-properties are the same. Some are more directly observable and public, like “walking,” while others are more private, like “thinking.” Nevertheless, if it were not possible to ascribe at least some instances of each type successfully from both a first- and a third-person perspective, it would not be possible to form and apply concepts of P-predicates at all. Thus, applying these concepts requires being able to apply them equally to self and other. This is possible, Strawson claims, only if our notion of person is more primitive than our notions of mind or body, and this notion of person that applies equally to self and other provides the ground for ascription of psychological predicates to self and other.

⁴ Strawson, *Individuals*, 104. ⁵ Strawson, *Individuals*, 104–105.

In advancing this view, Strawson claims to be engaging in what he calls “descriptive metaphysics,” which requires an analysis of how our everyday world presents itself to us as humans. He is not trying to provide a solution to the epistemological problem of knowledge of other minds since he sees the problem itself as incoherent. He says that once this problem is made coherent by explaining correctly how our concepts of mind are even possible, it dissolves. He also claims that his remarks on our acquisition of knowledge of P-properties in self and other are not intended “as *a priori* genetic psychology,” though he provides several hints that we will discuss here on how developmental psychologists might proceed.⁶ He wants only to make intelligible the conceptual scheme that we actually have. His solution is to show how our concept of person requires that both material and mental properties be ascribed to the same concrete individual and that the meaning of mental properties ascribed to self and other are the same, though the grounds for ascribing these properties vary due to differences in our perspectives of self and other. In what follows, we agree with Strawson’s assessment of the problem and argue for a unified concept of person of a Strawsonian sort as the basis of commonsense psychology.

From Strawson to a developmental psychology of persons and selves

On Strawson’s account of our acquisition of mental concepts, our acquisition of mental concepts presupposes that we have a prior intuition that persons are different from other objects. This intuition is not yet the acquisition of the concept of person as individual objects to which material and psychological predicates apply. Nor does this intuition provide concepts of psychological predicates. However, Strawson does provide two hints as to how a genetic or developmental psychology might proceed from this basic intuition to ascriptions of psychological properties. The first of these is to point out that some psychological predicates that we apply to self and other are so simple that the distance between the grounds for first-person ascription and third-person ascription is small. In the case, for instance, of simple actions, like walking or writing, physical movements and psychological properties are intermingled and inseparable. When we see someone walking, we immediately perceive the intentionality of this goal-directed activity as being the same as our own intentionality in walking. So, the gap between self and other is almost negligible in simple actions limited in their goal or object directedness.

⁶ Strawson, *Individuals*, 109.

Strawson suggests that here is where we might begin to understand how we bridge the gap between our different ways of knowing self and other, thus making a first step toward understanding psychological concepts in the same way for self and other. We will consider shortly how this idea fits within a developmental psychology of psychological ascription.

Strawson's second hint is his suggestion that individual ascriptions of psychological, or intentional, terms to self and other are not the only ways in which we ascribe such terms. In the case of activities involving forms of joint action, such as football (aka soccer), where the individual participates not so much with respect to individual intentions, but with respect to intentions or purposes of the group, Strawson briefly defends the intelligibility of the notion of "group mind." Each participant, though varying in physical behavior, shares a common intention or goal with other members of the same team. One can safely say that there is one goal here, not multiple goals, and this goal is a property of the group, a "we" goal that is singular, but nevertheless shared by individuals. 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. In what follows, we will see how this suggestion cashes out in a developmental psychology of our understanding of self and other as persons, and also show how joint cooperative activity and shared goals may have played a significant role in the evolution of the concept of person.

These two ideas of Strawson seem to have guided some recent attempts by philosophers to provide a Strawsonian "person theory" account of development that is opposed to traditional theory of mind (ToM) approaches of ST and TT.⁷ What we wish to do here is to be more explicit about how our own intentional relations theory (IRT) relates to Strawson's views and, in particular, to these two ideas. The result will be an updated account of IRT, in light of more recent developments both empirical and theoretical in psychology as well as to developments in other disciplines, such as comparative psychology, evolutionary biology,

⁷ Dow, On the joint engagement of persons; and Seemann, Person perception. In his article, Dow cites our own work as representing a person theory approach to the development of understanding of mental phenomena of persons. Indeed, Strawson's view, along with Wittgenstein's, did inspire us in developing our theory; see endnote 2 in Barresi and Moore, Intentional relations and social understanding.

Table 7.1 *A multi-level framework of social understanding of intentional relations*

| Level | Type | Represents | Information |
|-------|---------------|---|---|
| 1 | Separate | Only self activity Only other activity | First-person current of self Third-person current of other |
| 2 | Interpersonal | Joint agent We-activity | First-person current of self and third-person current of other |
| 3 | Personal | Embodied agent – self – other | First-person imagined and third-person imagined First-person imagined and third-person current |
| 4 | Mental | Mental agent | First-person imagined and third-person imagined |

and philosophy. In the present section, we will focus on the Strawsonian aspects in our model of human development. In subsequent sections, we will focus on phylogenetic origins of person concepts, and the potential evolutionary source and primary function of our concept of person.

The basic question that we put forward in our original article on IRT is: How is it possible for organisms like us to represent intentional activities (actions, emotions/motivations, and epistemic activities) of self and other in the same way, when the information that we process involving our own object-directed activities (called *intentional relations*) is of a radically different kind from the information that we process of the activities of another individual?⁸ Our solution, which we supported by a review of both phylo- and ontogenetic phenomena, is that we can come to a common understanding of the intentional activities of self and other only by recognizing matches between our own actions and the actions of others. In certain shared activities with others, in which mutual imitation often occurs, we recognize that matched intentional relations directed at objects appear in two forms depending on whether viewed from a first- or third-person perspective – forms that can be united into a single unified concept of the activity.

In our model (Table 7.1), we distinguish four levels in understanding of intentional relations that are of particular importance. On the first level, there are representations that help organisms recognize intentional activities of others different from the representations that guide their

⁸ Barresi and Moore, Intentional relations and social understanding.

own activities. With respect to their own intentional relations, organisms process information primarily about objects toward which their activities are directed, with minimal, and primarily implicit, information about themselves. By contrast, the information that they process of other organisms, primarily those of their own species, is directed at the animate movements of these others, with less attention and understanding of the objects toward which these agents are directing their activities. Thus, at this level, representations of their own activities differ from their representations of others. We view most organisms other than humans as operating at this level. Though they may acquire an understanding of the activities of others that goes beyond mere behavior toward intentional understanding, their representations of their own and other individuals' activities do not unite and form a single unified conceptual representation that can be applied equally to self and other. Most organisms never develop a concept of self that is on a par with their concept of other organisms. On our account, this means that they cannot apply to their own intentional activities the conceptual representations that they apply to the activities of others; and conversely, any conceptual representation that they have of their own activities cannot be applied equally to the activities of others. This is not to deny that there may be low-level representations that they may apply fairly equally to self and other; but these representations cannot be at the level of the whole organism as engaged in intentional activity. This must be the case because they do not have a representation of themselves as a whole organism that is in the same form as their representations of other organisms.

At our second level, we focus on joint activities in which it is possible to match one's own to another organism's activities directed at objects. In humans, this begins to occur in infants at about 9 months, in triadic relations with adults, where they engage in joint activities of various sorts directed at a variety of objects and not merely engage in face-to-face interactions directed at each other. We see this kind of joint activity as the primary ground upon which the infant comes to form an understanding of intentional activity directed at objects from both a first- and third-person perspective. It is this kind of joint engagement in a common activity, where attention is paid both to the other actor as well as to the objects of engagement, which enables the formation of a concept of common intentional activity that is bidirectional with respect to the perspective of each agent.

Joint engagement of this sort provides the means for eventual understanding of self and other as persons. Moreover, it also provides the ground for entry into the uniquely human world, which is socially

constructed rather than a natural inheritance of the species.⁹ However, at this age the infant has no more than an intuition that there is something special about interacting with other humans compared to inanimate objects and other beings. There is some evidence to suggest that the beginnings of the concept of person originate at this time, at least in recognizing other humans as members of a single category as well as in the special manner in which these infants interact with members of that category.¹⁰ These infants are capable of perceiving similarities in body parts of self and other that can be matched in imitative actions, which can be used to bridge the gap between intentional actions of self and others. Yet, they are still a long way from having a concept of individual person that they can apply both to self and other to which can be ascribed Strawsonian M- and P-properties. However, the basis for eventually forming such a concept is laid down in these interactions of joint activity directed at objects.

From the ninth to the eighteenth month, the infant goes through a rapid series of changes in understanding of joint activities and eventually achieves an understanding (presumably with Strawsonian unified concepts) of certain forms of intentional relations as properties of individual agents or persons including self.¹¹ The 9-month-old's engagement in joint activities starts off primarily in a passive role in the context of an adult's scaffolding of the joint activity, but as time moves on the infant becomes active in maintaining the activity. The infant's initial knowledge of the joint activity is sufficiently rich to perceive that she and the other individual are engaged with each other in an interaction involving some other object, but not rich enough to think of herself and the other equally as persons with particular intentional relations directed toward each other and toward the object. In terms of shared experience of the joint activity, this must be built up as the infant learns to behave in a coordinated fashion in the context of the scaffolded activity. In some cases,

⁹ J. Barresi, Some boundary conditions on embodied agents sharing a common world. In I. Wachsmuth, M. Lenzen, and G. Knoblich (eds.) *Embodied Communication in Humans and Machines* (Oxford University Press, 2008) 29–52; M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll, Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28 (2005) 675–735.

¹⁰ L. Bonatti, E. Frot, R. Zangl, and M. Mehler, The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognitive Psychology* 44 (2002) 388–426.

¹¹ C. Moore, Representing intentional relations and acting intentionally in infancy: Current insights and open questions. In G. Knoblich, I. Thornton, M. Grosjean, and M. Shiffrar (eds.) *Human Body Perception from the Inside Out* (New York: Oxford University Press, 2006) 427–442; and C. Moore and J. Barresi, The construction of commonsense psychology in infancy. In P. Zelazo, M. Chandler, and E. Crone (eds.) *Developmental Social Cognitive Neuroscience* (New York: Psychology Press, 2009) 43–62.

this involves imitation of adult actions, or vice-versa, imitation by an adult of an infant's actions. Turn-taking in actions and roles may sometimes occur in these situations. Whether the joint activity is a spontaneous form of play, or some repeated form of a more structured activity like "clean up" of toys, the infant gradually begins to understand basic rules of the activity, which allow it to determine whether the shared activity is occurring in a regular fashion. It is at this point that the infant can appreciate the "normative" quality of the joint activity and engage more actively in maintaining the joint activity. She can anticipate what the adult will or should do next and even lead the adult to do it through various communications that have become part of the activity – such as pointing to a new object to pick up after a previous object has been put away. The development of these skills shifts the form of joint activity from one that is primarily controlled by the adult's scaffolding behavior to one that is more collaborative and negotiated.¹²

There is a gradual shift during this period of development, from an appreciation of the "we" aspect of the experience, with its bipolar forms of representation of common actions of self and other that can be matched and integrated, toward a capacity to perceive diversity in the activities of self and other. Unified concepts of intentional activity that at an earlier stage could be used only when applied to joint actions of self and other, now can be applied individually to self or other. This is achieved through imagined rather than active sharing. When not engaged in joint action with an adult, the infant imagines being involved in the activity that she now understands that the adult is engaged in; or vice-versa, can imagine the adult as involved with her in her own activity. This makes it possible eventually for the infant to conceive of the intentions of others from a first- as well as third-person point of view even when she is not involved in the activity, and also to represent her own intentions as if another person were also engaged in the activity and observing her performance. Thus, the first- and third-person aspects of the activity can now be unified in interpreting the action of individuals, as well as in conditions of joint action.

The outcome of these processes of development is the formation of a concept of person as an embodied agent that can engage in a variety of intentional relations that are object or goal directed. This concept of person and of a person's properties, both material and psychological, can be applied to both self and other. It appears as level 3 of our model in Table 7.1. We propose that at this level the toddler of around age 2 has a reflective capacity to represent self as a fully embodied agent

¹² Tomasello *et al.*, Understanding and sharing intentions.

engaging in the same kinds of intentional activities as others. Moreover, the child can also represent other individuals as experiencing from a first-person perspective the same experiences that the child has when performing these intentional activities. 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 he 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 he can imagine the first-person perspective of the other in complementary positions in any activity.¹³ 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 their activity as that of two agents engaged in intentional activities, where both agents and their independent roles are understood so as to allow the child to take on either role if that were required. At this time, the child can also show empathy for the other, not only with respect to his own view of the situation they are in, but also with respect to the other’s particular stable desires and interests, some of which have been formed in previous joint activity.¹⁴ Mutual imitation becomes a game to play between peers in acquiring a better knowledge of how particular intentional relations appear from a first- and third-person position.¹⁵ Moreover, negotiating roles and choosing the direction of collaborative activity and play becomes increasingly a part of interactions not only with adults, who continue to scaffold normative requirements of culturally constituted activities, but also with peers in creative joint activities.

In sum, an infant’s achieving the ability to ascribe some psychological predicates equally to self and other begins in face-to-face interpersonal interaction, which lays the foundation for the infant’s differentiating persons from other objects. It then builds up as an understanding of mental phenomena in general out of sharing experiences with others in joint

¹³ cf. J. Martin and A. Gillespie, Chapter 8, this volume.

¹⁴ M. Tomasello and H. Moll, The gap is social: Human shared intentionality and culture. In P. M. Kappeler and J. B. Silk (eds.) *Mind the Gap* (Berlin/Heidelberg: Springer Verlag, 2010) 331–349.

¹⁵ M. Nielsen, The imitative behaviour of children and chimpanzees: A window on the transmission of cultural traditions. *Primatologie* (in press).

activities. Thus, both joint activities and simple actions enter into an understanding of how what appears in one way for self and another way for other can be united in a single concept of the activity – accessible in different ways for self and other. Intentional activities that the infant can ascribe equally to self and other thus start out with simple actions in which objects are directly involved and then progress to more complex intentional relations, such as emotions directed at objects in the immediate environment of the infant and adult, and then to epistemic relations such as seeing an object or not based on direction of gaze and what can be seen from a particular position. Desires based on previous emotional expressions as well as knowledge based on previous experience can also be represented at this level. However, the psychological properties that can be ascribed by the toddler at level 3 of our model are not yet meta-representational intentional properties, which have content that can be true or false.

Level 4 of our model makes these latter meta-representations possible. Instead of the concept of person as an embodied agent that persists through time, this level 4 concept is of a person, who is a mental agent, with beliefs and desires that can change, but would persist if not revised due to encounters with new experiences that provide a rational basis for change. This kind of mental agent and the associated person concept, match closely the Lockean model of personal identity based on a continued consciousness that links the past, present, and future self through time. Because this level does not provide new insights into Strawson's particular approach to person concepts, but plays a crucial role in distinguishing our human concept of person from related concepts in other animals, we will postpone discussing this level more fully until later in this chapter.

The phylogenesis of persons and selves

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

One form of evidence indicating the possibility that an animal can represent itself and others in a similar way appears when it is able to recognize itself in a mirror, especially the capacity to recognize a mark placed on a body part hidden from normal vision and to respond to it when placed before the mirror. While this test does not guarantee a concept of itself equivalent to that of others, it does indicate, along with the use of mirrors

to recognize other individuals, that the animal can treat the reflections in a mirror of itself and another equally as indicating a current physical appearance of a particular individual. Perhaps surprisingly, there are few animals that can recognize their full bodies in a mirror, though many can recognize other individuals in the mirror and can use the mirror for other purposes.¹⁶

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).¹⁷ 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 of non-kin.¹⁸ There is also evidence to suggest that great apes and dolphins can imagine the visual viewpoint and knowledge of others.¹⁹ Taken together, these findings suggest that such animals have comparable conceptions of self and other, conceiving of each from both a first- and a third-person point of view.

A question that remains to be answered is whether these two views of self and other are integrated? If not integrated, then there may not be a single form of representation, but two separate forms, that can be applied to both self and other. Thus, instead of level 3 of our model in Table 7.1, these animals may be only at level 1, albeit a more sophisticated level 1 than that of other organisms.²⁰ Alternatively, they may operate in some domains of intentional understanding at level 3 (what Moore calls

¹⁶ This procedure was developed by G. G. Gallup Jr., Chimpanzees: Self recognition. *Science* 167 (1970) 86–87. It should be noted that infants only begin to recognize themselves in the mirror at 18 months, when they first show a variety of other evidence that they have concepts of person and self that they can apply equally both to self and other (level 3 of our model). This makes the test a particularly appealing one to evaluate whether other animals develop something like our concepts of person and self.

¹⁷ G. G. Gallup Jr., J. R. Anderson, and S. M. Platek, Self-recognition. In S. Gallagher (ed.) *The Oxford Handbook of Self* (Oxford University Press, 2010) 80–110.

¹⁸ F. B. M. de Waal, Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology* 59 (2008) 279–300.

¹⁹ J. Call and M. Tomasello, Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Science* 12 (2008) 187–192; B. Hare, From hominoid to hominid mind: What changed and why? *Annual Review of Anthropology* 40 (2011) 293–309; L. Marino, R. C. Connor, R. E. Fordyce, L. M. Herman, P. R. Hof *et al.*, Cetaceans have complex brains for complex cognition. *PLoS Biology* 5 (2007) 966–972.

²⁰ We have suggested that this may be the situation for autistic individuals. In effect, the two forms of representation mimic ST and TT, because of their first- and third-person forms. See Barresi and Moore, The neuroscience of social understanding.

“intentional islands”), but remain at level 1 in other domains.²¹ This may occur because of limited experiences of shared intentional relations at level 2, which are necessary for domain general understanding of intentional relations at level 3. We believe that this latter alternative is more likely to be the case, at least with respect to the chimpanzee.

Evidence that chimpanzees are at level 3 of our model comes from their capacity for novel imitation. They are not only able to imitate complex actions involving unique combinations of elements but they can also recognize when another individual is imitating them.²² This ability to imitate novel actions and recognize being imitated suggests that a common form of representation is involved for self and other, which may eventually take the form of whole body representations. Thus, diverse actions of others can be understood as the same as one’s own actions. This may be the basis for the ability of chimpanzees and other imitative species to recognize themselves in a mirror, because mirror representation “imitates” their whole body self-image so closely that by varying their behavior before the mirror and “testing the image,” they may be able to distinguish the image as a representation of themselves.

Imitation appears to provide a link between self and other that makes possible a form of representation of actions that is the same for self and other even at the level of whole body. It is possible that these representations for chimpanzees are still at level 1 of our model, particularly since there is little evidence that chimpanzees can succeed in role reversal tasks, which we take as an important test of an organism’s capacity to conceive of self and other in the same way. But there may be other reasons for this failure. Role reversal tasks that have been tested so far on chimpanzees have involved cooperative tasks and chimpanzees are not very cooperative.²³ So, it may be that they do not attend well to the activity of the other chimpanzee in these tasks and merely attend to their own role. Consequently, when reversal occurs, they need to learn the complementary role.

It may be that while chimpanzees have level 3 forms of representations, their skill in and motivation for using these representations is limited. This hypothesis fits within a larger picture, which suggests that while chimpanzees can imitate others and can imagine the other’s epistemic viewpoint, they rarely use this ability compared to humans and it is used mostly in situations that serve self-interest in competitive contexts. Abundant recent research shows that chimpanzee understanding

²¹ Moore, Representing intentional relations and acting intentionally in infancy.

²² Nielsen, The imitative behaviour of children and chimpanzees.

²³ Tomasello and Moll, The gap is social.

140 *John Barresi, Chris Moore, and Raymond Martin*

of intentional relations of others is almost always limited to competitive rather than cooperative contexts. In the absence of specific training, it is as if they are almost blind to possible cooperative intentions of others. Moreover, they don't try to elicit help from others, or try to achieve joint goals with others. Tomasello and his colleagues suggest that this is due to limited interest in "shared intentionality," the desire to collaborate with others for common goals.²⁴ This contrasts sharply with human children who engage in shared intentionality during their second year. This notion of shared intentionality has much in common with our notion of shared intentional relations, but rather than focusing on our level 2, it elaborates on processes that occur cooperatively at our level 3. At this level, the activities of self and other are both represented in the same way, as integrated intentional relations of agents who share intentions. At level 2, on the other hand, self and other are represented differently, and only the shared intentional relations are represented in a unified way. Although Tomasello and his colleagues do not make this distinction explicit in their work, they do refer to our account of our notion of shared intentional relations when elaborating their own view.

What we want to focus on here is how face-to-face and joint activity in early infancy is important in distinguishing human from chimpanzee level 3 activities. It appears that chimpanzees do not engage much in face-to-face interaction and may not engage in any joint or shared activity involving other objects. The few examples of shared activity involving objects, such as tools, are brief, and show little sustained activity either by the adult or the infant. This contrasts sharply with human infant-adult interactions represented in level 2 of our model as shared intentional relations, which provide the foundation for later advances in collaboration. A chimpanzee infant may learn to use a tool by observing its mother use a tool or sometimes by the mother's making the tool available to the infant. This is nothing like the rich joint activity found in adult-infant human joint engagements, which involve a large number of socially constituted games and other activities, where roles can change and actions are coordinated with each other. In the case of humans, this shared activity involves shared intentional relations and later shared intentionality, and makes possible future cooperation, collaboration and cultural learning.

If this interpretation is right, it appears that chimpanzees can understand self and other as embodied agents of actions, but that rather than view self and other with this level 3 concept most of the time, they rarely use it. Usually, they represent self and other in distinct modes, which is

²⁴ Tomasello *et al.*, Understanding and sharing intentions; Tomasello and Moll, The gap is social.

at level 1 of our model. In terms of the Strawsonian theory of person, the lack of interpersonal relations of a cooperative nature in joint actions may limit their ability to represent intentional relations of self and other using a single uniform concept. If shared intentional relations ground human understanding of ourselves and others as persons, then the limited bridge between self and other that we see in level 3 chimpanzee behavior may not be enough. In sum, chimpanzees do not have a concept of embodied agent that they can apply equally to self and other across a variety of intentional relations and this may be due to a rather limited degree of early joint activities between infant and adult, which stems from their fundamentally competitive orientation toward each other. There is reason to think that bonobos, our other closest ape relative, are more affiliative in nature and more capable of entering into shared intentional activities in cooperative tasks.²⁵ Cetaceans are also more cooperative and on several measures seem to show a more advanced understanding of self and other as equivalent, intentional agents than either of these great apes.²⁶ But, we will not go further into these comparisons here. Instead, we will now discuss more generally the role of cooperation.

Reciprocal altruism and the evolution of the human concept of person

How did our human conception of person evolve out of earlier primate concepts? We think that the simple answer to this question is reciprocal altruism, a behavior explained by the evolutionary biologist Robert Trivers.²⁷ If there were not some long-term functional expectation of equal treatment with respect to costs and benefits of cooperation and mutual aid, such activity, and the psychological mechanisms upon which they depend, would not occur. Thus, it was the need to engage in joint activities and mutual aid that extended over time with the same partners, where self and other are treated equally with respect to costs and benefits, which made the concept of person necessary. However, it is important to distinguish between reciprocal altruism in terms of its evolutionary function, which moves altruism beyond the realm of kinship and inclusive fitness to include non-relatives, from the psychological mechanisms

²⁵ Hare, From hominoid to hominid mind.

²⁶ Marino *et al.*, Cetaceans have complex brains for complex cognition; A. Pack and L. M. Herman, Dolphin social cognition and joint attention: Our current understanding. *Aquatic Mammals* 32 (2006) 443–460.

²⁷ R. L. Trivers, The evolution of reciprocal altruism. *Quarterly Review of Biology* 46 (1971) 35–57.

that create and maintain this ultimate function. There have been significant theoretical and empirical advances since Trivers published his theory, many of which interpret reciprocal altruism as a simple “tit-for-tat” exchange of favors between unrelated individuals, then propose advances beyond this simple mechanism.²⁸ However, it seems that most of these advances fall within the general domain of Trivers’ initial comprehensive vision of function and mechanisms, and we will treat them as such here. For instance, the notion of altruistic punishment of defectors was contained in Trivers’ idea of “moralistic aggression.” Although group selection as a potential source of within-group altruism was not anticipated, Trivers’ notion of indirect, normatively based, reciprocity within groups is a ground that makes group selection possible.

We believe that the conditions that gave a selective advantage to increasing levels of within-group cooperation among non-kin and to more complex forms of reciprocal altruism within the hominid line provided the evolutionary basis for our human concepts of person and self. Essential for these human concepts of self and person are the capacities both to *represent* and to *weigh* the points of view of self and other equally. In cooperative relationships, where joint intentional activities are involved and one must take into account another person’s point of view and weigh it equally or almost equally to one’s own, the ability to represent and weigh intentional relations of self and other equivalently becomes important.

According to Trivers’ theory, costs versus benefits for self and other must be calculated and weighed with respect to action in a manner that will maintain equity across conditions and through time. This applies both to joint activities where benefits might be distributed equally or proportionally based on individual costs, as well as to situations that can warrant individual acts of altruism for another. For both joint collaborative activities and altruistic favors, maintaining equity requires a common metric to calculate costs and benefits across individuals, and not only from one’s own point of view. Otherwise, one might be able to calculate and weigh the costs and benefits for oneself for various actions in the present, and possibly into the future, without being able to calculate, or weigh, the costs or benefits for others. In these circumstances, each individual might pursue what seems evidently in their own self interest, while missing out on opportunities for cooperation over time and exchange of

²⁸ R. Axelrod and W. D. Hamilton, The evolution of cooperation. *Science* 211 (1981) 1,390–1,396; H. Gintis, S. Bowles, R. Boyd, and E. Fehr, Explaining altruistic behavior in humans. *Evolution and Human Behavior* 24 (2003) 153–172; M. A. Nowak, Five rules for the evolution of cooperation. *Science* 314 (2006) 1,560–1,563.

favors that would result in greater, long-term benefits both for self and others.

In Trivers' original presentation of the theory, the ability of individuals to calculate and weigh costs and benefits across self and other in a common metric appears to be assumed rather than explained. This would not be a problem, so long as costs and benefits of exchange were immediately advantageous to both sides. While a theory of mind and concept of person that applies equally to self and other would not be required in such cases, it does appear to be necessary for reciprocal altruism in humans. In our case, the use of a common metric requires such a theory in order to judge not only how costs and benefits appear to ourselves, but also how they appear to others. Moreover the theory must make it possible to estimate relative costs and benefits for agents with purposes that vary in their temporal extension. The use of a theory of this sort is essential in our attempts to maintain equity in relative costs and benefits to self and others in reciprocal altruistic relationships, where the medium of exchange is highly variable and depends on diverse individual purposes. Although we certainly have a bias in our own favor in our calculation of costs and benefits to self and other, and in weighing them in our actions, which Trivers associates with "subtle cheating," such cheating can only work in a context where gross differences in costs and benefits are easily detected, but minor deviations are not. To do this, we need a theory of mind, and associated concept of person, that applies equally well to self and other, both at a time and across time.

Trivers himself realized not long after publishing his theory that it could account for our sense of fairness and justice.²⁹ He had already postulated in the theory that friendship, sympathy, gratitude, guilt, reparation, trust, moralistic aggression, and other social emotions may have evolved as psychological mechanisms that track cost/benefit ratios in generating and responding to altruistic and non-altruistic acts, but he didn't mention our sense of justice or fairness. However, it seems likely that human moral psychology, particularly our concern with equity, emerged as a result of the evolution of reciprocal altruism in hominid species. This moral psychology requires a concept of person that we can apply equally to self and others, where the perspectives of different individuals, including their motivations and actions, can be interpreted and weighed using a common metric.

²⁹ R. L. Trivers, *Natural Selection and Social Theory: Selected Papers of Robert L. Trivers. (Evolution and Cognition Series)* (Oxford University Press, 2002) 16–17; Reciprocal altruism 30 years later. In P. M. Kappeler and C. P. van Schaik (eds.) *Cooperation in Primates and Humans: Mechanisms and Evolution* (Berlin: Springer-Verlag, 2006) 67–83; and *Social Evolution* (Menlo Park: Benjamin/Cummings Publishing Company, 1985).

If one compares adult humans to great apes in their capacity for such a concept of person, large differences occur both in their understanding and weighing of representational perspectives of self and others, both at a time and across time. Chimpanzees live mainly in the present, and, although they live in social groups with unrelated individuals, their cooperative activities and affiliative attitudes tend to be kin based. Some reciprocal relations and altruistic acts occur with non-relatives, but most of their activities either ignore non-relatives or are more competitive than cooperative with them. Since their attempts to understand each other tend not to be for purposes involving cooperation toward common goals, opportunities both to share intentional relations with common cooperative goals as well as to represent them are limited. Thus, to the extent that their concept of embodied agent depends on shared activity it must be quite narrow. Even adult–infant and other kin interactions tend to be limited in shared activity, so their concepts of each other’s intentional relations in general are likely to be isolated from their concepts of their own activities and more at level 1 of our model in Table 7.1 than at level 3 where individual person concepts first appear. As noted in the previous section, what concepts they have that apply across self and other are likely to be imitation-based rather than through original shared activity, and used primarily in competitive rather than in cooperative contexts.

It is also the case that chimpanzees do not act with regard to a distant future or a distant past. In contrast, humans have a notion of self that is extended in time and they have the capacity to engage in mental time travel.³⁰ This ability to imagine past and future events from different points of view applies to both self and other, and can also be applied to different representational points of view in the present. These capacities are linked together in humans. Our understanding of false belief in others and changing beliefs in self, requires the same ability for imagining changes in perspective through time. Moreover, this capacity is linked to our ability to act based on future motives that may be in conflict with current motives.³¹

³⁰ C. Moore and K. Lemmon (eds.) *The Self in Time: Developmental Perspectives* (Hillsdale, NJ: Erlbaum, 2001); T. Suddendorf and M. C. Corballis, The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences* 30 (2007) 299–351.

³¹ J. Barresi, Extending self-consciousness into the future. In Moore and Lemmon, *The Self in Time*, 141–161; C. Moore, J. Barresi, and C. Thompson, The cognitive basis of future-oriented prosocial behavior. *Social Development* 7 (1998) 198–218; Moore and Lemmon, *The Self in Time*; C. Thompson, J. Barresi, and C. Moore, The development of future-oriented prudence and altruism in preschool children. *Cognitive Development* 12 (1997) 199–212.

Whereas chimpanzees can act now for goals that will be achieved in the future, they cannot do so by delaying gratification of any current desire. By contrast, a 4-year-old child can delay gratification of a current desire in favor of a desire it knows it will have in the future. The child can also remember past episodes with different motives and knowledge as distinct from those he or she has in the present. Thus, the 4-year-old is capable of representing changes in representational perspective for self over time. And, the child's capacity to do this is linked to a similar capacity to represent diversity in the perspectives of others. Thus, in congruence with the need of a common metric for representing diverse perspectives that is required for reciprocal altruism of a general sort to occur, humans have it at an early age, while chimpanzees – and apparently other animals – never acquire it.

It has been argued that only narrow forms of reciprocation occur in non-human cases because most animals lack the mental capacities necessary to deal with costs and benefits extended across time.³² But, the evidence also suggests that they do not take into account the different perspectives of self and other, when conceiving of costs and benefits, and weigh them both when making decisions.³³ Thus, recent experimental findings are more congruent with the present proposal that what is unique about human reciprocal altruism is its dependence on a uniform concept of person that can be applied equally to self and other, which involves a common metric for calculating costs and benefits across diverse perspectives, and which extends across time. With this system in place, relatively short-term collaboration as well as long-term relationships involving reciprocal altruism of various kinds become possible and a sense of justice becomes a medium by which to think and act with respect to those relationships. With a notion of justice in place, large groups of individuals that mutually recognize each other as persons will be disposed to act even toward strangers in congruence with what is considered fair treatment among persons within their group or culture. This sense of justice can even reach beyond one's culture to humanity at large.³⁴

³² J. R. Stevens and M. D. Hauser, Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Sciences* 8 (2004) 60–65.

³³ K. Jensen, B. Hare, J. Call, and M. Tomasello, What's in it for me? Self-regard precludes altruism and spite in chimpanzees. *Proceedings of the Royal Society: B* 273 (2006) 1,013–1,021; J. R. Greenberg, K. Hamann, F. Warneken, and M. Tomasello, Chimpanzee helping in collaborative and noncollaborative contexts. *Animal Behaviour* 80 (2010) 873–880.

³⁴ J. Barresi, On seeing our selves and others as persons. *New Ideas in Psychology*, [in press](#).

Conclusion

In this chapter, we have looked into the evolutionary origin of our human concept of person and its early development. We have provided a person theory account of the development of human understanding of mental phenomena of self and other. In this account, we stress, with Strawson, that the notion of person is primitive and necessary in order to ascribe mental states with equivalent meaning to self and other. In our evolutionary as well as developmental account, we also follow Strawson in stressing the important role that shared mental states have in acquiring reflective understanding of these states and applying them to individuals. In our account, the fundamental asymmetry between the representations that most organisms use for self and other limits their understanding of mental states. It was the need for mechanisms that would support temporally extended cooperation between non-kin that we claim led to the human notion of person. Without the notion of person and a means for representing mental phenomena of self and other in a common format, opportunities for complex forms of reciprocal altruism between unrelated group members would be impossible. With it, there may be no limit on how far cooperation among humans who are willing and able to recognize each other as persons can reach.