Available online at www.sciencedirect.com
SCIENCE @ DIRECT®
Consciousness and Cognition 12 (2003) 549–571
Consciousness and Cognition
www.elsevier.com/locate/concog

# The emergence of a shared action ontology: Building blocks for a theory☆

Thomas Metzinger[a,*] and Vittorio Gallese[b,*]

[a] *Department of Philosophy, Johannes Gutenberg-Universität, D-55099 Mainz, Germany*
[b] *Dipartimento di Neuroscienze, Sezione di Fisiologia, Università di Parma, Via Volturno 39, I-43100 Parma, Italy*

Received 27 February 2003

**Abstract**

To have an ontology is to interpret a world. In this paper we argue that the brain, viewed as a representational system aimed at interpreting our world, possesses an ontology too. It creates primitives and makes existence assumptions. It decomposes target space in a way that exhibits a certain invariance, which in turn is functionally significant. We will investigate which are the functional regularities guiding this decomposition process, by answering to the following questions: What are the explicit and implicit assumptions about the structure of reality, which at the same time shape the causal profile of the brain's motor output and its representational deep structure, in particular of the conscious mind arising from it (its "phenomenal output")? How do they constrain high-level phenomena like conscious experience, the emergence of a first-person perspective, or social cognition? By reviewing a series of neuroscientific results and integrating them with a wider philosophical perspective, we will emphasize the contribution the motor system makes to this process. As it will be shown, the motor system constructs goals, actions, and intending selves as basic constituents of the world it interprets. It does so by assigning a single, unified causal role to them. Empirical evidence demonstrates that the brain models movements and action goals in terms of multimodal representations of organism-object-relations. Under a representationalist analysis, this process can be conceived of as an internal, dynamic representation of the intentionality-relation itself. We will show how such a complex form of representational content, once it is in place, can later function as a functional building block for social cognition and for a more complex, consciously experienced representation of the first-person perspective as well.


☆ We thank Wolfgang Prinz for a number of very helpful comments to an earlier version of this paper.
* Corresponding authors. Fax: +49-6131-3925141.
*E-mail addresses:* metzinger@uni-mainz.de (T. Metzinger), vittorio.gallese@unipr.it (V. Gallese).

1053-8100/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S1053-8100(03)00072-2

## 1. Introduction

Actions in the external world can be experienced as such, recognized, and understood. Simultaneously, the intentional content correlated with them (i.e., their satisfaction conditions resp. their intended goal-state) is interpreted by the observer as playing a causal role in determining the behavior of the observed other individuals. From a first-person perspective, the dynamic social environment appears as populated by volitional agents capable to entertain, similarly to the observer, an intentional relation to the world. We experience other agents as *directed* at certain target states or objects. We are "intentionality-detectors": As human beings, we cannot only mentally build an "objective," third-person account of the behaviors constituting the events of our social world. Beyond phenomenally experiencing the objective nature of a witnessed action, we can also experience its goal-directedness or intentional character, similarly to when we experience ourselves as the willful conscious agents of an ongoing behavior.

In the present paper we will provide and integrate some empirical and conceptual building blocks for a theory of the emergence of a common ontology between members of a group. We will examine, from a third-person scientific perspective, the fundamentally relational character of actions in the world. In particular, we want to look at the "ontological commitments" the brain makes when representing actions and goals. It will be further shown that the brain builds an ontology, an internal model of reality, which—on a very fundamental level within its representational architecture—incorporates the relational character of inter-actions between organism and environment, and that this architecture can actually be traced at the microfunctional level implemented in the brain's neural networks. The same subpersonal ontology then guides organisms when they are epistemic agents in a social world: Interpersonal relations become meaningful in virtue of a *shared action ontology*.

An action ontology can only be shared and successfully used by two systems, if there is a sufficient degree of functional overlap between them, if they decompose target space in similar ways. We will posit that the cognitive development of social competence capitalizes upon such a shared ontology to trigger the timely onset of behaviors such as gaze following, shared attention, and mind reading, which will eventually give rise to a full-blown capacity to entertain mental accounts of the behavior and goal states of other agents. We will also propose that what makes humans special is the fact that their functional ontology is much richer in *socially individuated* goal representations and that their model of reality is not only rich and flexible, but that they can actively *expand* their own functional ontology by mentally ascribing *distal* goals to conspecifics.

Neuroscientific results discussing the functional properties of mirror neurons, a class of premotor neurons in the monkey brain, will be introduced. It will be proposed that mirror neurons can be conceived of as the dawning of what the equivalent matching systems in our human brains are the fully developed realization of: a fundamental and mostly unconscious representational structure capable to build a *shared* action ontology. However, in closing we will also provide an example for a late and high-level utilization of this structure: enabling beings like ourselves to mutually *acknowledge each other as persons*, and to consciously experience this very fact at the same time.

## 2. The neural underpinnings of social understanding

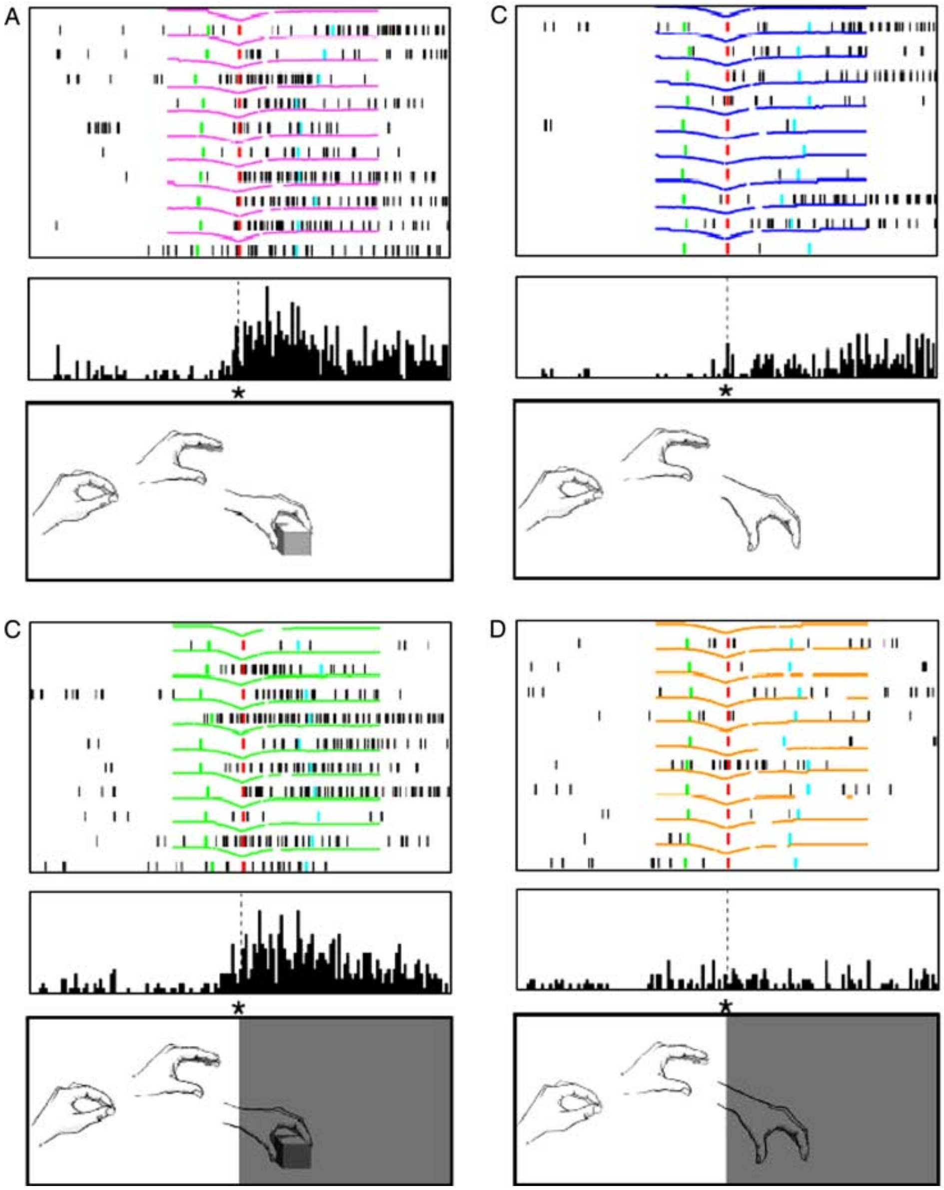
Primates, and particularly human beings are social animals whose cognitive development capitalizes upon the interaction with other conspecifics (adults, siblings, etc.). During social interactions we overtly manifest our inner intentions, dispositions and thoughts by means of overt behavior. We reciprocate this by trying to figure out what are the intentions, dispositions, and thoughts of others, when witnessing their behavior. Detecting another agent's intentions, or other inner states, helps anticipating this agent's future actions, which may be cooperative, non-cooperative, or even threatening. Accurate understanding and anticipation enable the observer to adjust his responses appropriately. Some recent neuroscientific results seem to suggest that a common neural representation underpins the specification of action end-states, independently of whose end-states are to be specified.

About 10 years ago a class of premotor neurons discharging not only when the monkey executed goal-related hand actions but also when observing other individuals (monkeys or humans) executing similar actions were discovered in the macaque monkey brain. These neurons were designated as "mirror neurons" (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Fogassi, & Gallese, 1996a; see also Fogassi & Gallese, 2002; Gallese, 2000, 2001; Gallese, Fogassi, Fadiga, & Rizzolatti, 2002; Rizzolatti, Fogassi, & Gallese, 2000; Rizzolatti, Fogassi, & Gallese, 2001; Rizzolatti, Craighero, & Fadiga, 2002). Mirror neurons require, in order to be activated by visual stimuli, an interaction between the agent (be it a human being or a monkey) and its target object. The visual presentation of objects does not evoke any response. Similarly, actions that, although achieving the same goal and looking similar to those performed by the experimenter's hand, are made with tools such as pliers or pincers have little effect on the response of mirror neurons (Gallese et al., 1996). Neurons with similar properties were later discovered in a sector of the posterior parietal cortex reciprocally connected with area F5, area PF or 7b (PF mirror neurons, see Gallese, Fadiga, Fogassi, & Rizzolatti, 2002b).

Of course, on the level of theories of mental representation, the idea of an "ideomotor principle" and the empirical hypothesis that perception and empathy engage motor representations are much older and go back at least to Howard Carpenter and William James. The discovery of mirror neurons, nevertheless, has changed our views on the neural mechanisms at the basis of interpersonal relations. It has been proposed that the mechanism instantiated by mirror neurons could be at the basis of an implicit form of action understanding (Gallese et al., 1996; Rizzolatti et al., 1996a). The observation of an action leads to the activation in the brain of the observer of the same neural network active during its actual execution: action observation causes action simulation, the automatic simulation of the motor plan leads to the same end-state in the observer. The shared, overlapping computational space leads to the implicit detection of the same end-state in the observed behavior of the agent (Gallese, 2003a; Gallese, Ferrari, & Umiltà, 2002a, 2002b).

The relationship between action understanding and action simulation is even more evident in the light of the results of two more recent studies. In the first series of experiments, Umiltà et al. (2001) tested F5 mirror neurons in two conditions: in the first condition the monkey could see the entire action (e.g., a hand grasping action); in the second condition, the same action was presented, but its final critical part, that is the hand-object interaction, was hidden. In the hidden condition the monkey only "knew" that the target object was present behind the occluder. The results showed that more than half of the recorded neurons responded also in the hidden condition (Umiltà et al., 2001) (see Fig. 1).

Behavioral data have shown that, like humans, monkeys can also infer the goal of an action even when the visual information about it is incomplete (Filion, Washburn, & Gulledge, 1996). The data by Umiltà et al. (2001) reveal the likely neural mechanism at the basis of this cognitive capacity. The inference about the goals of the behavior of others appears to be mediated by the activity of motor neurons coding the goal of the same action in the observer's brain. Out of sight is not "out of mind" just because, by simulating the action, the gap can be filled.



The results of a more recent series of neurophysiological experiments make this hypothesis even more plausible. Let us see why. Some transitive actions are characteristically accompanied by a sound. Imagine hearing the sound produced by footsteps of a walking person approaching you. This sound will induce you thinking that someone is getting closer to you. That particular sound enables you to understand what is going on, even if you have no visual information about what is currently happening out of your visual field. The action's sound has the capacity to make an invisible action inferred, and therefore present and understood.

To investigate the neural mechanism possibly underpinning this capacity, F5 mirror neurons were recorded from two monkeys during four different experimental conditions: when the monkey executed noisy actions (e.g. breaking peanuts, tearing sheets of paper apart, and the like); when the monkey saw and heard, or just saw, or just heard the same actions performed by another individual. The results showed that a consistent percentage of the tested mirror neurons fired when the monkey executed the action, just observed or just heard the same action performed by another agent (see Kohler et al., 2001, 2002; Keysers et al., 2003) (Fig. 2).

These neurons were defined as "audio–visual mirror neurons." They not only respond to the sound of actions, but also discriminate between the sounds of different actions. The actions whose sounds are preferred are also the actions producing the strongest responses when observed or executed. It does not matter at all for the activity of this neural network if the actions are specified at the motor, visual or auditory level. The activation of the premotor neural network controlling the execution of action A in the presence of sensory information related to the same action A can be characterized as simulating action A.

The multimodally driven simulation of action goals instantiated by neurons situated in the ventral pre-motor cortex of the monkey, exemplifies properties that are strikingly similar to the symbolic properties so characteristic of abstract human thought. The similarity with conceptual content is quite appealing: the same mental content ("the goal of action A") results from a multiplicity of states subsuming it, such as sounds, observed and executed actions. These states, in turn, are subsumed by differently triggered patterns of activation within a population of

Fig. 1. Example of an F5 mirror neuron responding to action observation in Full vision and in Hidden condition. The lower part of each panel illustrates schematically the experimenter's action as observed from the monkey's vantage point: the experimenter's hand starting from a fixed position, moving toward an object and grasping it (panels A and B), or mimicking grasping (panels C and D). The behavioral paradigm consisted of two basic conditions: Full vision condition (A) and Hidden condition (B). Two control conditions were also performed: Mimicking in full vision (C), and Mimicking hidden (D). In these last two conditions the monkey observed the same movements as in A and B, but without the target object. The black frame depicts the metallic frame interposed between the experimenter and the monkey in all conditions. In panels B and D the gray square inside the black frame represents the opaque sliding screen that prevented the monkey from seeing the experimenter's action performed behind it. The asterisk indicates the location of a marker on the frame. In hidden conditions the experimenter's hand started to disappear from the monkey's vision when crossing the marker position. The upper part of each panel shows rasters displays and histograms of ten consecutive trials recorded during the corresponding experimenter's hand movement illustrated in the lower part. Above each raster kinematics recordings (black traces) of the experimenter's hand are shown. The black trace indicates the experimenter's hand movements recorded using a motion analysis system. The illustrated neuron responded to the observation of grasping and holding in Full vision (A) and in the Hidden condition (B), in which the interaction between the experimenter's hand and the object occurred behind the opaque screen. The neuron response was virtually absent in the two conditions in which the observed action was mimed (C and D). Histograms bin width = 20 ms. Ordinates: spikes/s; abscissae: time (modified from Umiltà et al., 2001).

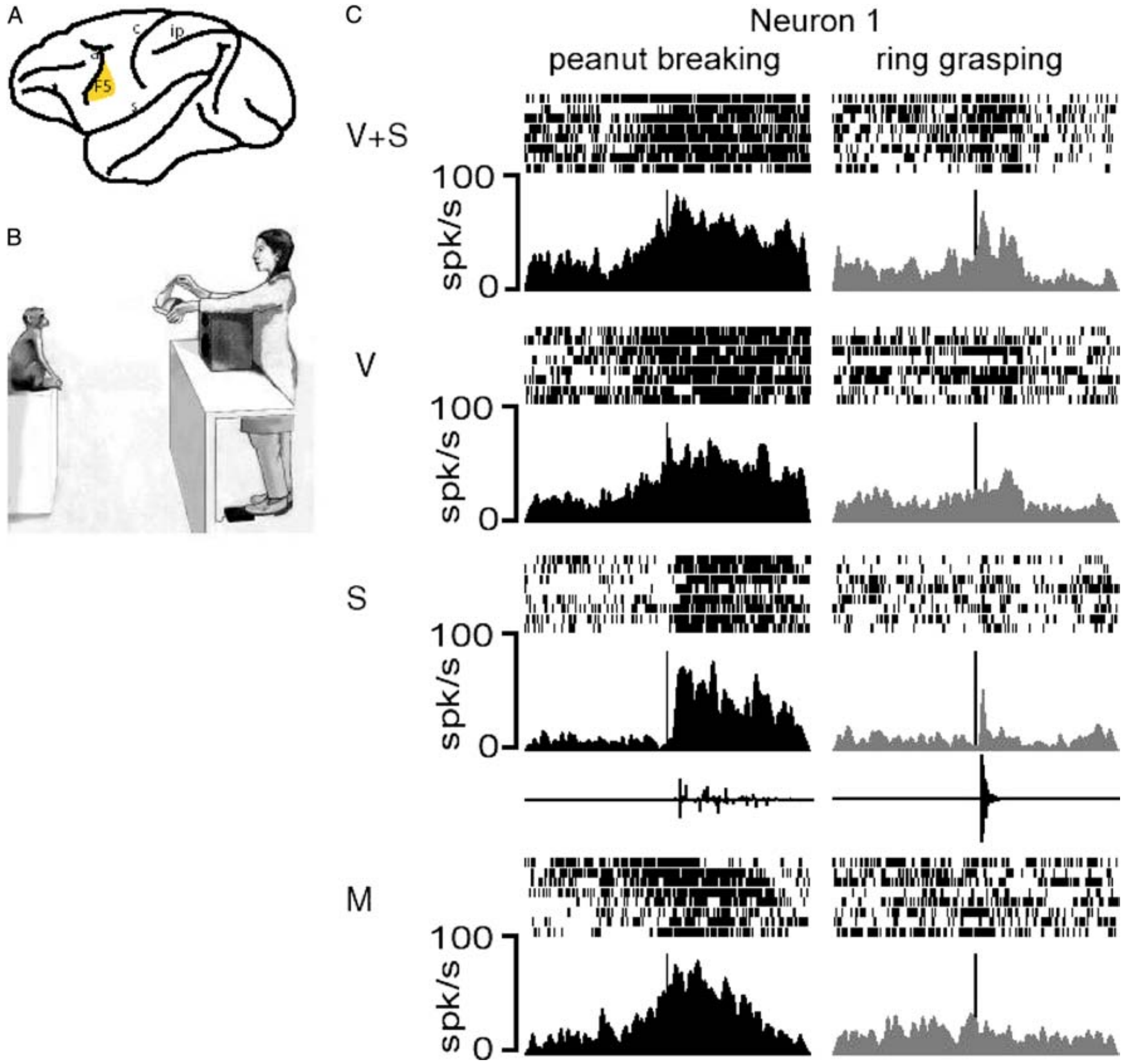



Fig. 2. Example of an F5 audio–visual mirror neuron. (A) Lateral view of the macaque brain with the location of area F5 shaded in gray. Major sulci: a, arcuate; c, central, ip, intraparietal; s, sylvian sulcus. (B) Experimental setup. (C) Response of a neuron discriminating between two actions in all modalities. Rastergrams are shown together with spike density functions for the best (black) and the less effective action (gray). V + S, V, S, and M stand for vision-and-sound, vision-only, sound-only, and Motor conditions, respectively. The vertical lines indicate the time at which the sound occurred (V + S, S) or would have occurred (V).The traces under the spike-density functions in the sound-only conditions are oscillograms of the sounds played back to test the neurons. This neuron discharged when the monkey broke a peanut (row 'M') and when the monkey observed the experimenter making the same action (rows V and V + S). The same neuron also responded when the monkey only heard the sound of a peanut being broken without seeing the action (row 'S'). When the monkey grasped a ring ('M'), Neuron 1 responded much less, demonstrating the motor specificity of the neuron. Also both the vision and the sound of an experimenter grasping the ring determined much smaller responses. A statistical criterion yielded both auditory and visual selectivity for this neuron (modified from Keysers et al., 2003).

"audio–visual mirror neurons." Depending on one's theory of concepts, all this may of course also be interpreted as a process of *non-conceptual* generalization. The important point, however, is that we have evidence for an abstract, allocentric, and multimodal level of goal representation (see Gallese, 2003b). The general picture conveyed by these results is that the premotor-parietal F5-PF mirror matching system instantiates simulations of actions utilized not only to generate and control goal-related behaviors, but also to provide a meaningful account of the goals and purposes of others' actions, by means of their implicit and automatic simulation.

What is the relation between these data and our understanding of human social cognition? Several studies using different experimental methodologies and techniques have demonstrated in humans also the existence of a similar mirror system, matching action observation and execution (see Buccino et al., 2001; Cochin, Barthelemy, Lejeune, Roux, & Martineau, 1998; Decety et al., 1997; Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Grafton, Arbib, Fadiga, & Rizzolatti, 1996; Hari et al., 1998; Iacoboni et al., 1999; Rizzolatti et al., 1996b). In particular, it is interesting to note that brain imaging experiments in humans have shown that during action observation there is a strong activation of premotor and parietal areas, the likely human homologue of the monkey areas in which mirror neurons were originally described (Buccino et al., 2001; Decety & Grèzes, 1999; Decety et al., 1997; Grafton et al., 1996; Iacoboni et al., 1999; Rizzolatti et al., 1996b). From an empirical point of view, it is now highly plausible to assume that in standard configurations for humans, as well as for monkeys, action observation automatically triggers action simulation on a rather abstract, multimodal level of neural representations. Before moving upwards from the neuroscientific and functional to the representationalist level of analysis, let us define more precisely what we imply by "simulation."

## 3. Embodied simulation

So far we have characterized the function of mirror neurons in terms of implicit action understanding, by means of embodied simulation. The notion of simulation is at present employed in many different domains, often with different, not necessarily overlapping meanings. Simulation is a functional process that possesses a certain representational content, typically focussing on the *temporal evolution* or on *possible* states of its target object. For example, an authoritative view on motor control characterizes simulation as the mechanism employed by forward models to predict the sensory consequences of impending actions (see Wolpert, Doya, & Kawato, 2003). According to this view, the predicted consequences are the simulated ones. In philosophy of mind, on the other hand, the notion of simulation has been used by the proponents of Simulation Theory of mind reading to characterize the pretend state adopted by the attributer in order to understand others' behaviour (see Goldman, 1989, 1992; Gordon, 1986, 1995).

We will use the term "simulation" as the core element of an *automatic*, *unconscious*, and *pre-reflexive control functional mechanism*, whose function is the modelling of objects, events, and other agents to be controlled. Simulation, as conceived of in the present paper, is therefore not necessarily the result of a willed and conscious cognitive effort, aimed at interpreting the intentions hidden in the overt behavior of others, but rather a basic functional mechanism of our brain. However, because it also generates representational content, this functional mechanism seems to play a major role in our epistemic approach to the world. It represents the outcome of a *possible* action one could take, and serves to attribute this outcome to another organism as a *real* goal-state it tries to bring about. Indeed, successful perception requires the capacity of predicting upcoming sensory events. Similarly, successful action requires the capacity of predicting the expected consequences of action. As suggested by an impressive and coherent amount of neuroscientific data (see Gallese, 2003a), both types of predictions seem to depend on the results of unconscious and automatic simulation processes. In this sense, simulation is not simply confined to the domain of motor control, but it seems to be a more general and basic endowment of our brain. It is *mental* because it has a content, but it can be *motor*, because its function may be realized by the motor system. We also call it "embodied"—not only because it is neurally realized, but also because it uses a pre-existing body-model in the brain, and therefore involves a prerational form of self-representation. In this context, low-level action simulation in social cognition can also be seen as an *exaptation*: There has never been any "special design" for the function we describe here. It is an extended functionality that was later co-opted from a distinct original adaptational functionality (namely, motor control), and it did not arise as a consequence of progressive adaptations via natural selection, because the underlying anatomical structure was never *directly* selected for.

A further point that deserves some comments is the relationship between a given observed action and the simulation process it induces. How can it be possible that when witnessing a given action, *exactly the same action* is simulated in the observer's brain? How is action identity internally represented (see Knoblich, this issue). How can the appropriate simulation be computed? Two hypothetical scenarios can be sketched. According to the first hypothesis, the selection process is determined on the basis of *prior visual identification* of the observed action. Once the action is visually identified, its simulation can ensue. On the basis of this scenario, though, the simulation process appear to be redundant for understanding what the observed action really means, because it implies its prior visual identification. The capacity of mirror neurons to be activated by hidden actions, provided that their outcome can be predicted (see above, Umiltà et al., 2001), seems to suggest, however, that action simulation does not require a comprehensive visual representation of the witnessed action in order to be triggered. According to the second hypothesis—the one we endorse—the selection process of which action is to be simulated in response to the observed one is automatic and pre-determined (see the "direct matching hypothesis," Rizzolatti et al., 2001). Now, the crucial point is to explain what enables the direct matching between the visual representation of an action and its motor representation.

The answer to such a question can only be speculative. On the one hand, early forms of automatic imitation of mouth movements in newborn babies (Meltzoff & Moore, 1977) seem to suggest a hard-wired mechanism coupling action observation with action execution. Perhaps it is no coincidence that these phenomena are mostly involving the action of body parts, such as the mouth, for which we have no direct visual control. In the case of hand actions, the ontogenetic development of the direct matching mechanism could instead be the result of a Hebbian association between the agent's own action and self-observation, and a further generalization of this association to others.

## 4. Sharing an action ontology: the general idea

One useful way to look at the brain is to describe it as a dynamic representational system. Every representational system presupposes an ontology: A set of assumptions about what the elementary building blocks of the reality to be represented actually are. By necessity, it constructs primitives. For example, many natural languages, viewed as representational systems, typically assume that extralinguistic reality is constituted by objects, properties, and relations. Their underlying ontology then frequently carries over into folk-psychological contexts, for instance, by influencing the way in which we naively describe our own mental states in everyday life situations. A quantum mechanical description of reality in physics, on the other hand, may successfully operate under a completely different set of metaphysical assumptions. It represents reality under a different ontology, because it uses different primitives. An interesting, but frequently overlooked fact is that the human brain possesses an ontology too, because it makes assumptions about what the relevant and what the irreducible building blocks of external reality are.

By metaphorically speaking of the brain as "making assumptions" (or as "arriving at conclusions," etc., see below), of course, we do not mean to imply that the brain is a cognitive or epistemic agent, or even that it internally uses propositional formats and symbolic coding principles. *Persons* make assumptions, and brains are only subpersonal functional modules of whole persons. In addition, the representational dynamics of brains defies the distinction between syntax and semantics. It has to be conceived as an "agent-free" type of subpersonal self-organization. Personal-level predicates, or what philosophers sometimes call "intentionalist idioms," have to be avoided on all subpersonal levels of description. There is no little man in the head interpreting quasi-linguistic representations, and, for the approach sketched here, the phenomenal self is something that has no transcendental subject or conscious agent "behind it". However, in generating a coherent internal world-model, brains decompose target space in a certain way. The functional regularities guiding this decomposition process are by no means arbitrary. We are interested in the causally relevant invariance guiding it, and this is what we mean by investigating the brain's "ontological assumptions." Please note that we will here take no position at all on the ontological commitments of cognitive neuroscience or the metaphysics generated by scientific theories in general.

Furthermore, although we are interested in how this process constrains and enables high-level phenomena like subjectivity and phenomenal experience, we do not want to imply a solution to the problem of consciousness as such. As will become obvious, the representational dynamics we describe cuts across the border between phenomenal and non-phenomenal states in an interesting way. Our main point is that there exist unconscious functional precursors of what can later also be phenomenally represented as a goal, an acting self or an individual first-person perspective. This is a point where some conceptual tools provided by philosophy of mind can shed more light on the general problem.

Two research targets are of particular interest in this context, because they enrich the system's functional ontology in a decisive way: The phenomenal self-model (PSM) and the phenomenal model of the intentionality-relation (PMIR). The PSM is the *integrated* conscious model an organism may have of itself as a whole. The PMIR is the integrated conscious model an organism may have of itself as standing in a relation to certain perceptual objects, other agents, and, in particular, the *goals* of other agents. In folk-psychological terms, the content of the PSM is what we sometimes naively call "the" self. The content of the PMIR is what we call "our" perspective on the world and other subjects (see Metzinger, 2003 for details). Our claim is that the mirror system makes a centrally relevant microfunctional contribution to both of these high-level phenomena, because, on the neural level, it helps to *code* the representational content in question:

First, the ongoing activity of the motor system obviously is a major contributor to any system's current self-model; second, there now is strong empirical evidence for mirror neurons as specifically coding organism-object relations on various levels of abstraction. We will come back to this point at the end of our paper. Before we can do so, it will be necessary to further clarify our conceptual framework and show how it is integrated with the data upon which we draw.

## 5. Goals

We think that the neuroscientific data so far presented make it necessary to conceptually restructure the questions that have to be asked. Let us begin by defining some basic notions. What is a goal? From a strict scientific point of view, no such things as goals exist in the objective world. All that exists are goal-*representations*, for instance, as we have seen in Section 2, those activated by biological nervous systems. A goal-representation is, first, formed by the representation of a certain state of the organism, of the world, or by the holding of a certain relation between the organism and a part of the world, e.g., another organism. Goal-representations are representations of goal-states. Second, what functionally *makes* such a state a goal-state is the fact that its internal representation is structured along an axis of valence: It possesses a value *for* the system. Broadly speaking, a value is anything that is conducive to preserving an organism's integrity (e.g., homeostasis), to maintaining integration on higher levels of complexity (e.g., cognitive development and social interaction), and to procreative success. Therefore, the reward system will be a second important element of the way in which a goal-representation can be implemented in a causally effective way. Goal-states imply values on the level of the individual organism, and values are made causally effective through the reward system.

It is interesting to note how infants differently construe goal-relatedness when witnessing the intentional actions of other individuals as opposed to physical events not involving human agents. When 18-month-old infants see a person slip and fail to complete an intended action, they imitate the intended action and not the actual movements that the actor made. However, if the action is displayed by a mechanical device, they fail to successfully reproduce it (Meltzoff, 1995). A further argument favoring the hypothesis that goal-relatedness is differently perceived by infants in social and physical event configurations, is provided by some recent findings by Amanda Woodward and collaborators (Woodward et al., 2001). These researchers have shown that 6-months-old infants react differently to observed grasping actions according to the biological (human hand) or artificial (mechanical claw) nature of the grasping agent. Only the former are considered as goal-directed actions. It appears therefore that infants' early propensity to attend to goals seems to be specific to human actors.

Biological systems are systems under evolutionary pressure, systems having to predict future events that possess a high survival value. Therefore, the prehistory of representational goal-states is likely to be found in the reward system. Why is this so? Reward is the payoff of the self-organizing principles that functionally govern and internally model the organization of an open system such as a living body is. Every organism has an internal, likely predetermined and genetically imprinted "drive" pushing toward homeostasis. A reward system is necessary to tell the organism that it is doing right, that it is heading on along the right track, the track leading through multiple stages to the achievement of higher and more complex levels of integration.

Higher integration means greater flexibility, which in turn means fitness, better adaptation to changing environments, better chances to pass over genes, and the like. In healthy individuals, the architecture of their goal-state hierarchy expresses their individual "logic of survival" (Damasio, 1999).

Therefore, the reward system can be conceived of as generating a new, non-symbolic kind of representational content, namely the internal value assigned to a certain state. It is important to note that even if "value" (just like "goal") in the normative sense is not a public property observable from a scientific third-person perspective, an internal simulation of value within an individual organism can be causally effective, and adaptive. A conscious representation of value, as for instance expressed in a subjectively experienced emotional state, has the additional functional advantage of making survival value-related information globally available for the selective and flexible control of action, attention, and cognition within a virtual window of presence. Global availability means that this information is accessible to many different processing systems at the same time (Baars, 1988; Metzinger, 2003). Presence means that it is functionally integrated into short-term memory and therefore phenomenally experienced as holding now.

A further interesting aspect of the relation between goal-state representations and reward consists in the fact that very soon the infant also learns to rely more and more on *external* causes for the internal activation of the reward system. The positive reactions (or their lack) induced in the adult caregivers by infant's behavior provide her almost on-line with a very useful guide about how to act in a given context. It has indeed been shown that around six months of age infants visually "check back" to the mother's emotional reaction in order to disambiguate ambiguous or uncertain events. Such phenomenon has been called "social referencing" (see Thompson, 1999). Once the value of the adults' reaction has been evaluated, a given goal-state representation can reliably be consolidated (or put into question, in case of negative reactions), making it a stable part of the infant's behavioral repertoire.

Neurally realized goal-state representations possess a number of interesting features. First, they change the system's functional ontology: It now acts as if something like goals actually were a part of reality, as if maximizing complexity and integration was a good in itself. As a representational system, it embodies an assumption, and this assumption—a representational construction of valence—becomes causally active. To use a metaphor coined by Francisco Varela, teleology is now "enacted" (Varela, Thompson, & Rosch, 1991).

Second, on the level of conscious goal-representation, goal representations frequently are phenomenally transparent. This is to say that the organism "looks through" the actual representational mechanism and "directly" assigns value to perceived target objects, to specific subject–object-relations characterizing certain goal-states, or even to another agent. Phenomenal transparency turns an organism into a naive realist. Because earlier processing stages are not accessible to introspective attention, their content appears as directly and immediately given. On the level of conscious experience, the internal construct of "valence" therefore frequently becomes an objective property perceived in external parts of reality. However, at least in humans, phenomenally opaque goal-representations do exist as well: In these cases we consciously experience them *as* representations. We experience ourselves as beings that actively *construct* their own goals, and then operate with internal representations of the corresponding goal-states. Conscious goal representations are positioned on a continuum between transparency and opacity, with a strong bias towards transparency (for more on phenomenal transparency, see Metzinger, 2003, Sections 3.2.7 & 6.4.2).

A third feature makes goal-representations interesting: Goal-representations do not possess truth-conditions, but conditions of satisfaction. Because no such things as goals exist in the objective order of things, a goal-representation cannot be true or false. However, the goal-states as represented can either hold or not hold, or they can hold to various degrees. Therefore, an internally represented goal-state can continuously be matched against sensory input (or even against ongoing internal simulations like memories, or the results of planning operations).

Fourth, goal-representations are intimately linked to an organism's self-representation: In standard situations they typically depict an individual logic of survival. Only on higher levels of complexity do we find mentally represented first-person plural and socially individuated goal-states. As representations are also dynamical entities, it is best to conceive of the special form of representational content formed by goal-states as an ongoing process that allows an organism to functionally appropriate the fact that certain states of itself, of the world and its social environment are valuable states to itself.

The possession of goal-states, if integrated into a coherent self-model, allows an organism to own the logic of its own survival—functionally, representationally, and sometimes even consciously. Goal-states are states the organism wants to bring about, by means of goal-representations. Goals—as representational constructs—are fundamental elements of the brain's model of the world. In particular they are building blocks of behavioral space (as represented by the brain). They turn its functional ontology into a teleological ontology, or, in short, into a "teleontology."

On the other hand, the building blocks of behavioral space seem to be abstract representations of possible actions, as seen from no particular perspective, because the empirical evidence (e.g., on mirror neurons, see above) now clearly seems to point to the existence of agent-independent goal detectors. Therefore, a particularly interesting way of analyzing the content of goal-representations is as states that portray *a successful, completed action from no particular/individual perspective*. The new theoretical problem to be solved is to explain how precisely these allocentric entities later get bound to the conscious perspective of a full-blown agent. But what is an "action"? And what is a "perspective"?

## 6. Actions

On a non-conceptual level, actions are elementary building blocks of reality for certain living organisms: Some kinds of organisms have developed agent-detecting modules, and some of them also conceive of themselves as agents. They have an extended ontology, because their reality has been considerably enriched. We can define such functionally enriched systems as possessors of an "action ontology." Let us now define what an action is on a *conceptual* level. Let us begin by distinguishing movements, behaviors, and actions.

Bodily movements are simple physical events, and they can be represented accordingly. Behaviors are movements that are goal-directed, i.e., which can meaningfully be described as directed towards a set of satisfaction conditions, but without necessarily being linked to an explicit and conscious representation of such conditions. As simple motor acts, they also do not have a consciously experienced reward-producing component (Rizzolatti et al., 2001, p. 668). In particular, behavior is something that can take place in a semi-automatic fashion, like when brushing our teeth.

Actions are a specific subset of goal-directed movements: A series of movements that are functionally integrated with a currently active representation of a goal-state as leading to a reward constitute an action. Therefore, an action is not isomorphic to a particular movement or specific behavioral pattern, because many different movements can constitute the same goal-directed action. What individuates an action is the set of satisfaction conditions defining the representational content of its goal-component as leading to a reward plus the special way in which it is causally linked to the actual event of overt movement generation. In particular, an action results from a selection process (which may or may not be conscious) and a representation of the system as a whole as standing in a certain relation to a specific goal-state (which is phenomenally represented, e.g., globally available via short-term memory).

The second defining characteristic is that an action in the true sense not only involves an explicit and conscious self-representation, but also a representation of the perspective the system now takes onto the world. That is, the selection process may well be unconscious, but it inevitably leads to a more global final stage resulting in a conscious representation of the system as a whole—as *having* an intention, as initiating and executing its own bodily movements. In other words, on the phenomenal level we always find a corresponding global state in which the system as a whole is itself represented as an agent.

This leads us to the third, and most intriguing, aspect of actions in the strong sense we are here proposing: Actions are first-person phenomena, and they are carried out by a conscious self. It is important to understand that all of this does not necessarily involve reflexive self-consciousness, the possession of concepts, or the mastering of a language: In animals such as monkeys an attentional and a volitional perspective could suffice to establish the first-person character of actions. In order to be appropriately related to an action goal it is enough to be able to (non-cognitively, sub-symbolically) attend to it or to (non-cognitively, sub-symbolically) select a specific motor pattern. At least these are two further assumptions underlying the way in which the brain, as a representational system, typically decomposes the reality it models into sub-components. For full-blown actions, we can only understand the special causal linkage between a goal-representation and the overt motor output, if we also understand how both are mediated through a specific representational structure, namely the "phenomenal model of the intentionality relation" (PMIR, see above). This structure will be explored in the next section.

## 7. The PMIR: A short representationalist analysis

### *7.1. What is a PMIR?*

Before addressing the problem of the phenomenal aspect of the shared action ontology, we must first focus on the way the agent-object relationship can be phenomenally represented. What is the phenomenal model of the intentionality relation (for details see Metzinger, 2003, Chapter 6)? It is a conscious mental model, and its content is an ongoing, episodic subject–object-relation. Here are four different examples, in terms of typical phenomenological descriptions of the class of phenomenal states at issue: "I am someone, who is currently visually attending to the color of the book in my hands," "I am someone currently grasping the content of the sentence I am reading," "I am someone currently hearing the sound of the refrigerator behind me," "I am someone now deciding to get up and get some more juice."

The central defining characteristic of phenomenal models of the intentionality-relation is that they depict a certain relationship as currently holding between the system, as transparently represented to itself, and an object-component. Phenomenologically, a PMIR creates the experience of a self in the act of knowing or of a self in the act of intending and acting. In the latter case, a goal-representation is transiently integrated into an ongoing action-simulation. This class of phenomenal mental models is particularly rich, because the number of possible object components is almost infinitely large.

Phenomenal models of the intentionality-relation typically consist of a transparent subject component and varying object components. Those can be transparent as well as opaque, transiently being integrated into an over-arching, comprehensive representation of the system as standing in a specific relation to a certain part of the world. The overall picture emerging is that of the human self-model continuously integrating the mechanisms of attentional, cognitive, and volitional availability against a stable background, which is formed by the transparent representation of the bodily self (Metzinger, 2000, 2003, Chapter 6). If one now takes the step from a representationalist level of description to the actual phenomenological changes inherent in the emergence of a full-blown conscious first-person perspective, it is easy to see how for the first time it allows a system to consciously experience itself as being not only a part of the world, but of being fully *immersed* in it through a dense network of causal, perceptual, cognitive, attentional, and agentive relations.

### *7.2. The phenomenal representation of volition*

Given the notion of a PMIR, we can begin to develop a more fine-grained analysis for the phenomenological target properties of volitional subjectivity and agency. Conscious volition is generated by integrating abstract goal representations—constituted by allocentric motor-simulations—into the current model of the intentionality-relation as object-components, in a process of decision or selection. Let us differentiate a number of cases. If we *contemplate* a certain action goal, i.e., when we ask ourselves whether we should get up and walk over to the refrigerator, we experience ourselves as cognitive subjects. This kind of phenomenally represented subject–object relationship can be analyzed as one in which the object-component is phenomenally opaque: Experientially, we know that we take a certain attitude toward a self-generated *representation* of a goal. A completely different situation ensues if we integrate a goal representation into the phenomenally transparent self-model, thereby making it a part of ourselves by *identifying* with it. Obviously, goal representations and goal hierarchies can also be important components of self-models that are not based on transient subject–object relations, but on enduring internal reorganizations of the self-model, of its emotional and motivational structure, etc., and which may possibly last for a lifetime.

A volitional first-person perspective—the phenomenal experience of practical intentionality—emerges if two conditions are satisfied. First, the object-component must be constituted by a particular self-simulatum, by a neural simulation of a concrete behavioral pattern, e.g., like getting up and walking toward the refrigerator. Second, the relationship depicted on the level of conscious experience is one of currently *selecting* this particular behavioral pattern, as simulated.

Again, we can usefully illustrate this by describing it as "representational identification": The moment following volition, the moment at which concrete bodily behavior actually ensues, is the moment in which the already active motor simulation is integrated into the currently active bodily self-model, and thereby causally coupled to the rest of the motor system and the effectors. It is precisely the moment, in which we *identify* with a particular action, transforming it from a possible into an actual pattern of behavior and thereby functionally as well as phenomenologically embodying it. Embodiment leads to enaction. Interestingly, the moment of agency seems to be the moment when the phenomenal model of the intentionality relation collapses. We can now describe the experience of being a volitional subject and the experience of being an agent more precisely, using the simple tools already introduced.

Phenomenal volition is a form of conscious content. It can be analyzed as representational content as follows: [I myself ( = the currently active transparent model of the self) am currently ( = the *de-nunc*-character of the overall phenomenal model of the intentionality relation, as integrated into the virtual window of presence) present in a world ( = the transparent, global model of reality currently active) and I am just about to select ( = the type of relation depicted in the phenomenal model of the intentionality relation) a possible way to walk around the chairs toward the refrigerator ( = the object-component, constituted by an opaque simulation of a possible motor pattern in an egocentric frame of reference)].

The experience of agency follows in the moment in which the internal "distance" created between phenomenal self-representation and phenomenal self-simulation in the previously mentioned structure collapses to zero: I *realize* a possible self-state, by enacting it. As I experience myself walking around the chairs and toward the refrigerator, proprioceptive and kinesthetic feedback allows me to feel the degrees to which I have already identified with the sequence of bodily movements I have selected in the previous moment.

Please recall how phenomenally transparent representations are precisely those representations the existence of whose content we cannot doubt. They are those we experience as real, whereas opaque representations are those, which we experience as thoughts, as imagination, or as hallucinations. Realizing a simulated self-state means developing a functional strategy of making it the content of a transparent self-model, of a self that really exists—on the level of phenomenal experience.

*Ongoing* agency, the conscious experience of sustained executive control, can therefore be representationally analyzed according to the following pattern: [I myself (the content of the transparent self-model) am currently ( = the *de-nunc* -character of the phenomenal model of the intentionality relationship as integrated into the virtual window of presence) present in a world ( = the transparent, global model of reality) and I am currently experiencing myself as carrying out ( = continuously integrating into the transparent self-model) an action which I have previously imagined and selected (the opaque self-simulation forming the object-component, which is now step by step assimilated into the subject-component)]. Of course, there can be all sorts of additional functional and representational complications, e.g., if the proprioceptive and kinesthetic feedback integrated into the internal model of the body does not match with the forward model still held active in working memory.

It is interesting to see how agency conceived of as executive consciousness ("Vollzugsbewusstsein" in the sense of Karl Jaspers) can be analyzed as an ongoing representational dynamics collapsing a phenomenal model of the practical intentionality relationship into a new transparent self-model. Again, as the whole structure is embedded into a virtual window of presence, the transparent, intranscendable experiential state for the system as a whole now is one of being a full-blown volitional subject, currently being present in a world, and acting in it.

Finally, the PMIR has a phenomenally experienced direction: PMIRs are like arrows pointing from self-model to object-component. As soon as one has understood the arrow-like nature of the PMIR, two special cases can be much more clearly described. First, the arrow can point not only outwards, but also *inwards*, namely in cases where the object component is formed by the PSM itself (as in introspective attention or in thinking about oneself). Here, the second-order PMIR internally models a system–system-relationship instead of a system–object-relationship. Furthermore, in consciously experienced *social* cognition the object-component can now be either formed by a phenomenal model of another agent or an arrow *in* the other agent's head (as in observing another human being observing another human being). Such ideas are appealing, because they show how the relevant representational domain is an open domain: In principle, many layers of complexity and intersubjective metacognition can be added through a process of social/psychological evolution. As the elementary representational building block, the PMIR, gets richer and more abstract, an ascending and cognitively continuous development from the simple portrayal of body-centered subject–object-relations to full-blown self-other modeling becomes conceivable. What changes from monkeys to humans is only the complexity of the self-modeling process.

## 8. Intentionality phenomenalized

The philosophical step just taken consists in *phenomenalizing intentionality*. Phenomenalizing intentionality, we submit, may be an indispensable first step in the project of naturalizing intentionality tout court. Meaning and the conscious experience of meaningfulness have to be separated. Of course, this does not mean that no such thing as intentional content exists.

Mental representations possess at least two kinds of content: Phenomenal content (PC) and intentional content (IC). As a large majority of researchers in philosophy of mind currently agrees, PC can be characterized by the Principle of Local Supervenience (PLS):

> **(PLS):** By nomological necessity, once all internal and contemporaneous properties of an organism's nervous system are fixed, the PC of its mental states is fixed as well.

Very obviously, the same is not true of IC: What a representational state is a representation *of* also depends on the environment of the organism, e.g., on *what* it actually represents and if it does so correctly or adequately. The PC, on the other hand, is something a veridical perception and a hallucination may share. The phenomenal content of your mental representations is that aspect which, being independent of their veridicality, is available for conscious experience from the first-person perspective, while simultaneously being determined by inclusively internal properties of your brain. (Metzinger, 2000, p. 3; for more on supervenience, see Kim, 1993). However, what is not yet determined is if this experiential character also goes along with actual knowledge about the world, with intentionality in terms of a biological information-processing system generating states possessing representational—or intentional—content. IC, in many cases, is determined by external and non-local factors. PC, on the other hand, is what stays invariant regardless of whether a perception is veridical or a hallucination.

The frequently overlooked point, to which we here draw attention, is that intentionality (on a pre-rational level, probably starting with the sensori-motor system and early levels of attentional processing) is *itself* depicted on the level of phenomenal content. And it is precisely this kind of conscious content, which has guided theoreticians for centuries in developing their own, now classic theories of intentionality. Due to the principle of local supervenience (see above), it has today become highly plausible that this aspect of intentionality can be naturalized. The phenomenal experience of being an intentional agent, of being a perceiving, attending and cognizing subject can be naturalized. Of course, this does in no way preclude the possibility that intentional content as such can never, and maybe even for principled reasons, be naturalized. However, getting the first obstacle out of the way may greatly help in gaining a fresh access to intentionality as such, because it frees us from the burden of false intuitions generated by our own transparent model of reality and because it helps us to set aside the issue of how we come to consciously experience our mental states as meaningful and directed towards an object-component. We can separate the issue of consciously experienced intentionality from the more general problem of how something like representational content could evolve in the minds of human beings and other animals *at all*.

In this context, it is interesting to note how the originally philosophical concept of a conscious model of the intentionality-relationship (Metzinger, 1993, 2000, 2003) currently surfaces at a number of places in the cognitive neurosciences. Jean Delacour, in an excellent review of current ideas about possible neural correlates of conscious experience, explicitly introduces the notion of an "intentionality-modeling structure" (Delacour, 1997, p. 138). LaBerge (1997, p. 150, p. 172) elucidates how important an understanding of the self-representational component present in attentional processing will have to be for a full blown theory of conscious attention. Craik and colleagues (1999) point out how episodic memory, of course, is a process of reconstructing what was here termed a PMIR, because one necessary constituent of memory retrieval is not simply the simulation of a past event, but an association of this simulation with a self-representation (Craik et al., 1999, p. 26). Building an autobiographic memory is a process of self-related encoding, and conscious, episodic memory retrieval is a process necessarily involving the self-model, because reactivating a PMIR inevitably means reactivating a PSM. Most notable, of course, is Antonio Damasio's conception of a "juxtaposition" of self and object (see Damasio & Damasio, 1996a, 1996b, p. 172, p. 24) and the general framework of a fully embodied "self in the act of knowing" (Damasio, 1994, 1999).

The theory we are trying to build is at present mute about the question if anything like "real" intentionality does exist. A possible speculation is that philosophical models of the intentionality of the mental have ultimately resulted from a naive realistic interpretation of the process of visual attention, of the phenomenal self directing its gaze at a visual object, thereby making it more salient, and simply elevating this interpretation to the level of epistemology. The concept of the intentionality of the mental may simply be a mistaken attempt to theoretically model epistemic relations in accordance with the consciously experienced model of the intentionality relation. Pursuing this point is outside the scope of the present article. However, we will briefly point to two interesting aspects of the overall issue, which are of considerable theoretical interest.

Of particular interest is the fact that the brain models the relationship between subject and object as an asymmetric relationship. It is the consciously experienced "arrow of intentionality," paradigmatically experienced in having the feeling of "projecting" visual attention outwards, as it were, or in attentionally "tracking" objects in the environment. *Intendere arcum*, to bend the bow of the mind and point the arrow of knowledge toward parts of the world is an intuitively plausible and popular philosophical metaphor, in particular in combination with the idea of "direct," magical intentionality.

We can now understand why such an idea will strike beings like us as intuitively plausible: It is *phenomenally* possible, because there is a directly corresponding structural element in our conscious model of reality. Many theoretical models of the representational relationship are implicitly oriented at the phenomenal experience of visual attention, of the directedness inherent in the phenomenal model of the intentionality relation. Frequently, the theoretical model we design about ourselves as cognitive agents is one of organisms, which, *ad libitum*, direct the beam of their "epistemic flashlight" at parts of the world or their own internal lives, of beings, which *generate* the representational relation as subjects of experience. This can lead to the kind of fallacy, which Daniel Dennett has described as "Cartesian materialism" (Dennett, 1991, p. 333). As Dennett has pointed out, many of the different forms of Cartesian materialism, the assumption of a final inner stage, can also be generated in the context of representationalist theories of mind by mistakenly transporting what he called the "intentional stance" (Dennett, 1987a) into the system, (see Dennett, 1991, p. 458). The model here proposed, of course, does not make this mistake; it is much more closely related to an extension of the idea of a "second-order intentional system" (Dennett, 1987b), a system that applies the intentional stance to itself—but in a phenomenally transparent manner.

A related hypothesis is that philosophical theorizing about the intentionality relation has generally been influenced by that aspect of our phenomenal model of reality, which is generated by our strongest sensory modality. If the process of mental representation in general is modeled in accordance with our dominant sensory modality (namely, vision), we will automatically generate distal objects—just as we do in our transparent, visual model of reality. If the object-component of a PMIR is of an opaque nature, as in genuinely cognitive contents or in goal representation, a philosophical interpretation of these mental contents as non-physical, "intentionally inexistent" objects in the sense of Brentano (1874) becomes inevitable.

## 9. Consciously sharing an action ontology: A phenomenological multi-level analysis

Phenomenal mental models are instruments used to make a certain subset of information currently active in the system globally available for the control of action, for focal attention and for cognitive processing. A phenomenal model of transient subject–object relations makes an enormous amount of new information available for the system: All information related to the fact that it is currently perturbed by perceptual objects, that certain cognitive states are currently occurring in itself, e.g., to the fact that certain abstract goal representations are currently active, that there are a number of concrete self-simulations connecting the current system-state with the state the system would have if this goal state would be realized; allowing for selective behavior and the information that it is a system capable of manipulating its own sensory input, e.g., by turning its head and directing its gaze to a specific visual object. A first-person perspective allows a system to conceive of itself as being part of an independent objective order, while at the same time being anchored in it and able to act on it as a subject (see, e.g., Grush, 2000).

Let us now move to the social dimension. Once a system is capable of representing transient subject–object relations in a globally available manner it becomes possible for the object-component in the underlying representational structure to be formed by the intentions of other beings. Once again the brain's ontology is expanded, and, as we have learned in the preceding sections, the motor system plays a crucial role in this functional expansion: A phenomenal first-person perspective allows for the mental representation of a phenomenal second-person perspective. Therefore, it is of central relevance to find the subpersonal building blocks implementing this capacity.

A representationalist analysis of the matching mechanism instantiated by mirror neurons clearly shows that the object component of an other-agent PMIR simulated by F5 mirror neurons does not have to be visible in order for the full-blown neural response to occur (Umiltà et al., 2001). Mirror neurons code not object-presence, but rather the *relational* fact that a certain external agent is currently directed at an object-component. The natural default assumption, therefore, is that if the other agent is conscious they code the existence of a PMIR in another agent, and interestingly they do so in virtue of being a central part of the system that, in other situations, constructs an internal PMIR in the monkey herself.

To have a PMIR inevitably means to have a phenomenal self-model. But to be self-conscious does not imply having language, concepts, being able to mentally form a concept of oneself, etc. Body image and visceral feelings are enough. Because the monkey's motor system allows for prediction of actions/occluded target objects, neurons responsive to the observation of goal-related behaviors might actually be triggered not by a visual representation alone, but by the "embedded motor schema", that is the same motor schema which is at work when the monkey itself is acting in a similar way.

We posit that the representational deep structure of this schema is what later evolved into the full-blown PMIR in human beings. It should, again, be stressed that the notion of a PMIR as here introduced does not yet imply of being able to mentally form a concept of oneself *as* oneself, of consciously experiencing the selection process for goal-states, etc. An elementary self-model in terms of body image and visceral feelings plus the existence of a low-level attentional mechanism is quite enough to establish the basic representation of a dynamic subject–object relation. The non-cognitive PMIR is thus what builds the bridge into the social dimension. Once a rudimentary subjective perspective has been established with the help of the motor system, inter-subjectivity can follow.

The dynamics of low-level intersubjectivity then helps to further develop, enrich, and stabilize the individual first-person perspective in each participating agent. If a functional mechanism for discovering and phenomenally representing the unobservable goal states of conspecifics is in place, the observed behavior of other systems in the organism's environment can lead to the activation of a goal-representation, which in turn can be represented as belonging to someone else. This would be a decisive enrichment of the organism's functional ontology, because they would then irrevocably become a part of it's reality, leading to a change in its behavioral profile. As we have seen, it is empirically plausible that such a mechanism is actually in place in human beings. Therefore, representations of the intentions of external agents can now become the object-component of the phenomenal model of the intentionality-relation as well. This enables further high-level phenomena.

Behavior-reading is transformed into mind-reading, because the representational tools for action representation allow for action simulation as well (see Gallese, 2003a, 2003b; Gallese & Goldman, 1998), including the simulation of goal-states (i.e., of specific, successfully terminated actions, of new and successfully achieved subject–object-relations) in other agents. If this happens on the level of conscious experience, a completely new and highly interesting form of information is integrated into a virtual window of presence and becomes globally available for the system: The information of actually *now* standing in certain relations to the goals of other conspecifics. We would claim that it is precisely the conscious availability of this type of information, which turned human beings from acting, attending and thinking selves into social subjects.

In this paper we have mainly been concerned with the investigation of subpersonal elements of the brain's "action ontology." Let us close with a very specific high-level capacity that could arise from such elements. If the fact that your are constantly not only standing in perceptual and behavioral relations to your environment, but that you are frequently realizing subject–subject-relationships becomes globally available, it also becomes available for cognition. This, in turn, will allow those systems capable of concept formation to cognitively model social relations from a third-person perspective. Such beings can mentally represent social relationships between other individuals depicted as intentional agents, even if they are not involved themselves. Second, if such systems possess or develop the concept of a "person", they can now mutually apply it to each other. Through our extended PSMs we were able to simultaneously establish *correlated* cognitive PMIRs of the type just described. Two or more human beings could now *at the same time* activate cognitive PMIRs mutually pointing to each other under a representation as rational individuals. And the correlated nature of these two mental events, their mutuality and interdependence, could itself be represented on the level global availability. In this way we were able to mutually *acknowledge each other as persons*, and to consciously experience this very fact. We will not pursue this point at length here, but it is obvious how this particular representational capacity is of high relevance for social cognition and the pooling of cognitive resources. Social cognition, empathy, and cognition in general, need a minimal level of complexity in the brain's functional ontology, and in human beings and primates the premotor cortex seems to be substantial part of the physical substrate underpinning this level.

## Acknowledgments

Vittorio Gallese was supported by MIURST and the Eurocores Program of the European Science Foundation.

Thomas Metzinger was supported by the McDonnell Project in Philosophy and Neuroscience.

## References

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Brentano, F. (1973) [1874]. *Psychologie vom empirischen Standpunkt*. Erster Band, Hamburg: Meiner (English Trans.: Brentano, F. [1874](1973). *Psychology from an Empirical Standpoint*. Edited by Oskar Kraus. English edition edited by Linda McAlister. Translated by Antos, C. Rancurello, D.B. Terrell and Linda McAlister. London: Routledge & Kegan Paul/New York: Humanities Press).

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., & Freund, -. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience, 13*, 400–404.

Cochin, S., Barthelemy, C., Lejeune, B., Roux, S., & Martineau, J. (1998). Perception of motion and qEEG activity in human adults. *Electroencephalography and Clinical Neurophysiology, 107*, 287–295.

Craik, F. I. M., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., & Kapur, S. (1999). In search of the self: A positron emision tomography study. *Psychological Science, 10*, 26–34.

Damasio, A. R. (1994). *Descartes' error*. New York: Putnam/Grosset.

Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt, Brace and Jovanovich.

Damasio, A. R., & Damasio, H. (1996a). Images and subjectivity: Neurobiological trials and tribulations. In R. N. McCauley (Ed.), *The churchlands and their critics*. Cambridge, MA: Blackwell.

Damasio, A., & Damasio, H. (1996b). Making images and creating subjectivity. In R. Llinás, & S. Churchland (Eds.), *The mind–brain continuum*. Cambridge, MA: MIT Press.

Decety, J., & Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences, 3*, 172–178.

Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F., & Fazio, F. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain, 120*, 1763–1777.

Delacour, J. (1997). Neurobiology of consciousness: An overview. *Behavioural Brain Research, 85*, 127–141.

Dennett, D. C. (1987a). *The intentional stance*. Cambridge, MA: MIT Press.

Dennett, D. C. (1987b). Intentional systems in cognitive ethology: The Panglossian Paradigm defended. In: Dennett 1987a. First published in *The Brain and Behavioral Sciences,* 6, 343–390.

Dennett, D. C. (1991). *Consciousness explained*. Boston, Toronto and London: Little Brown.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology, 73*, 2608–2611.

Filion, C. M., Washburn, D. A., & Gulledge, J. P. (1996). Can monkeys (Macaca mulatta) represent invisible displacement? *Journal of Comparative Psychology, 110*, 386–395.

Fogassi, L., & Gallese, V. (2002). The neural correlates of action understanding in non human primates. In M. Stamenov, & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 13–36). Amsterdam, Philadelphia, PA: John Benjamins.

Gallese, V. (2000). The acting subject: Towards the neural basis of social cognition. In T. Metzinger (Ed.), *Neural correlates of consciousness. Empirical and conceptual questions* (pp. 325–333). Cambridge, MA: MIT Press.

Gallese, V. (2001). The "Shared Manifold" Hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies, 8*(5–7), 33–50.

Gallese, V. (2003a). The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society London B, 358*, 517–528.

Gallese, V. (2003b). A neuroscientific grasp of concepts: From control to representation. *Philosophical Transactions of the Royal Society London B, 358*, 1231.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 12*, 493–501.

Gallese, V., Ferrari, P. F., & Umiltà, M. A. (2002a). The mirror matching system: A shared manifold for intersubjectivity. *Behavioral Brain Sciences, 25*, 35–36.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593–609.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (2002b). Action representation and the inferior pariatel lobule. In W. Prinz, & B. Hommel (Eds.), *Common mechanisms in perception and action: Attention and Performance* (Vol. XIX, pp. 334–355). Oxford: Oxford University Press.

Gallese, V., Fogassi, L., Fadiga, L., & Rizzolatti, G. (2002). Action representation and the inferior parietal lobule. In W. Prinz, & B. Hommel (Eds.), *Attention and performance XIX. Common mechanisms in perception and action* (pp. 334–355). Oxford: Oxford University Press.

Goldman, A. (1989). Interpretation psychologized. *Mind and Language, 4*, 161–185.

Goldman, A. (1992). In defense of the simulation theory. *Mind and Language, 7*, 104–119.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind and Language, 1*, 158–171.

Gordon, R. M. (1995). Simulation without introspection or inference from me to you. In M. Davies, & T. Stone (Eds.), *Mental simulation* (pp. 53–67). Oxford: Blackwell.

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by PET: 2. Observation compared with imagination. *Experimental Brain Research, 112*, 103–111.

Grush, R. (2000). Self, world and space: The meaning and mechanisms of ego- and allocentric spatial representation. *Brain and Mind, 1*, 59–92.

Hari, R., Forss, N., Avikainen, S., Kirveskari, S., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 15061–15065.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*, 2526–2528.

Keysers, C., Kohler, E., Umiltà, M. A., Fogassi, L., Nanetti, L. & Gallese, V. (2003). Audio-visual mirror neurones and action recognition. *Experimental brain research*, in press.

Kim, J. (1993). *Supervenience and mind*. Cambridge: Cambridge University Press.

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297*, 846–848.

Kohler, E., Umiltà, M. A., Keysers, C., Gallese, V., Fogassi, L., & Rizzolatti, G. (2001). Auditory mirror neurons in the ventral premotor cortex of the monkey. *Social Neuroscience Abstract*, XXVII(129), 9.

LaBerge, D. (1997). Attention, awareness, and the triangular circuit. *Consciousness and Cognition, 6*, 149–181.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838–850.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*, 75–78.

Metzinger, T. (1993;[2]1999). *Subjekt und Selbstmodell*. Paderborn: Mentis.

Metzinger, T. (2000). The subjectivity of subjective experience: A representationalist analysis of the first-person perspective. In T. Metzinger (Ed.), *Neural correlates of consciousness—Empirical and conceptual questions*. Cambridge, MA: MIT Press.

Metzinger, T. (2003). *Being no one. The self-model theory of subjectivity*. Cambridge, MA: MIT Press.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2000). Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (2nd ed., pp. 539–552). Cambridge, MA: MIT Press.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2, 661–670.

Rizzolatti, G., Craighero, L., & Fadiga, L. (2002). The mirror system in humans. In M. Stamenov, & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 37–62). Amsterdam, Philadelphia, PA: John Benjamins.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1996a). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., & Fazio, F. (1996b). Localization of grasp representation in humans by PET: 1. Observation versus execution. *Experimental Brain Research, 111*, 246–252.

Thompson, R. A. (1999). Empathy and its origins in early development. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 145–157). Cambridge: Cambridge University Press.

Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing: A neurophysiological study. *Neuron, 32*, 91–101.

Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind*. Cambridge, MA/London, UK: MIT Press.
Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society London B, 358*, 593–602.