Available online at www.sciencedirect.com
SCIENCE DIRECT®
Consciousness and Cognition 12 (2003) 487–494
Consciousness and Cognition
www.elsevier.com/locate/concog

Introduction

Grounding the self in action

Günther Knoblich,[a,*] Birgit Elsner,[b] Gisa Aschersleben,[a] and Thomas Metzinger[c]

[a] *Günther Knoblich, Max-Planck-Institute for Psychological Research, Amalienstrasse 33, 80799 Munich, Germany*
[b] *Universität Heidelberg Germany*
[c] *Johannes Gutenberg-Universität Mainz, Germany*

Received 18 August 2003

Consciousness and cognition are phenomena that seem to be inextricably bound to an individual first-person perspective: at least in standard situations, there is not only conscious experience, but also an experiencing *self*. And there is not only thought as such, but a thinking *self* as well. Why is there not only the flow of experience, but also *someone*—someone who *has* these experiences? And why do most thoughts not just occur in a free-floating way, like clouds in the sky, but seem to originate from—and within—a *thinking self*, a self somehow mentally portrayed as an independent cause in itself, a cognitive *agent*?

Presently, in a number of different disciplines, two general answers seem to slowly emerge. First, the subject of consciousness and the subject of thought are frequently present in the unfolding process of phenomenal experience and cognition. They are present in the form and content of a *self-representation*, i.e., a representation of the experiencing, thinking system as a whole, which can be more or less explicit, more or less accessible to introspection, and more or less conscious. This self-representation is central to any understanding of the deep structure of mind. Second, the theoretically most relevant aspect of phenomenal and cognitive self-representation is likely to be found in the aspect of *agency*. It has to do with action control. The experience of agency is of maximal importance in constituting a first-person perspective as well as in the development of empathy and social cognition. This is so, because selfhood is not only characterized by phenomenal and representational properties, but also by highly specific *functional properties*—properties that somehow turn biological organisms into agents, sometimes even into persons.

In order to fully understand these functional properties we need to know more about how they are implemented: neurobiologically, behaviorally, and socially. And we need a truly

* Corresponding author. Fax: +49-89-38602-199.
E-mail address: knoblich@psy.mpg.de (G. Knoblich).

1053-8100/$ - see front matter
doi:10.1016/j.concog.2003.08.009

multidisciplinary approach, integrating areas as diverse as the philosophy of mind, cognitive, developmental, and social psychology, psychiatry, and neuroscience. This special issue tries to make a bold first step by drawing a number of different and usually unconnected threads together, thereby arranging a number of traditional lines of research in a new way, deliberately putting them into focus in a slightly new manner. In doing this, we also hope to draw attention to a new field of investigation that, in our view, will have a major impact on future research addressing consciousness and cognition.

Two important developments currently guide the research directed at the self. First, it became increasingly clear that defining the self as a central and unitary processing module, or as a central representation in a knowledge network, is hardly compatible with the enterprise of naturalizing the mind. On the other hand, taking phenomenology seriously, one has to admit that there exists a *phenomenal self*, typically characterized by just these properties of centrality and unity, which are so hard to find on functional and representational levels of description. Second, an increasing number of researchers started to believe that it is impossible to understand the self without grounding it in action. In other words, the self needs to be embedded into the causal network of the physical world. Thus, today, there is not only a "symbol grounding problem" (Harnad, 1990), but also a "self grounding problem." Although specific disciplines have tried to tackle this problem for quite some time, only recently a new multidisciplinary perspective on the self has begun to emerge. Currently, the self grounding problem is the focus and the frontier in this new line of research.

This special issue aims at providing a state-of-the-art account of this new perspective. Some controversial issues are discussed in commentaries and the authors' responses. The contributions are arranged according to five broad research areas, all of which have explored the significance of action for being a self and becoming a self. The first section addresses different conceptual frameworks that aim at grounding the self in action. (Proust; Prinz; Newen & Vogeley; Metzinger & Gallese). The second section addresses the cognitive and neural systems that underlie action identification, including the ability to distinguish between self-produced and other-produced actions (Decety & Chaminade; Leube et al.; Farrer et al.; Knoblich & Flach; Jordan). The third section is concerned with the role of action for explaining disorders of the self (Blakemore, Daprati et al., Kircher & Leube). The fourth section explores the role of volition and intention in shaping the experience of oneself, and the awareness of one's own actions (Wegner & Erskine; Haggard & Clark; Wohlschläger et al.). Finally, the fifth section addresses the role of action in self-development (Rochat; Elsner & Aschersleben; Kiraly et al.; Sodian et al.), which is an important issue in developmental psychology. In the following, we provide an overview of the articles in each of these five sections.

(1) Conceptual foundations. The first set of contributions tries to lay some conceptual foundations for the general enterprise of a unified theory of self and action. The first two articles are of a purely theoretical nature, whereas the second two are each co-authored by a philosopher and a neuroscientist, leading on to the more empirical parts of this special issue. The last paper of this section also constitutes a thematic bridge to the following section, which focuses on self-other relationships.

Joëlle Proust's opening piece analyzes the conditions under which a person is constituted by being able to recognize herself *as herself* across different points in time. The philosophical problem of personal identity consists in large part in discovering a way of defining a notion of self

(to which the target property of "ipseity" then corresponds), which, as Proust writes, "allows understanding how an individual can be—and represent herself as—the same self, although her mental and bodily dispositions vary considerably, as well as the environment in which she is leading her life." After discussing the shortcomings of some traditional approaches Proust proposes an architecture in which overlapping metacognitive memory is combined with *mental agency*: A self emerges if a system has started to simultaneously monitor and to actively control the content of its own mental states, i.e., of its knowledge, desires, emotions, or attentional processes; and if it is then able to generate an *integrated* representation of itself as having successfully revised and controlled its own mental states in the past, and of itself as being able to do so right now. Joëlle Proust closes her discussion by applying her model of "ipseity" as metacognitive memory specialized in dynamic belief/desire revision to certain pathologies of the self and competing approaches to their theoretical interpretation.

Wolfgang Prinz adds a hypothetical evolutionary scenario to this discussion. According to the scenario, mental selves emerged, because at a certain stage biological organisms had to solve a source-attribution problem. For Prinz, "self-morphic" modes of representation can only emerge if two developmental stages follow each other in sequence: At the first stage, it is necessary to develop the ability to re-present circumstances that are not currently present and keep such representation separate from perception. He calls such combined online/offline-capacities dual representation. At the second stage—named attribution to persons—these must be integrated with an interpretation of some representations as resulting from personal communication. The emergence of a mental self solved a central problem that now occurred: thoughts as internally generated acts of re-presentation are not accompanied by the perception of any current act of communication—so they cannot be attributed to any external human source in the current situation. It is interesting to note, therefore, how the hypothetical account of Wolfgang Prinz spans two major levels, because dual representation concerns the natural history of behavior organization; attribution to persons, in contrast, concerns the cultural history of our species. One important and provocative implication of this model is the way in which subjectivity is construed outward-in, that is, one's own mental self is derived from, and is secondary to, the mental selves perceived in others. In conclusion Prinz also points out one major consequence of this approach—namely how the social construction of subjectivity and selfhood relies on, and is maintained in, various types of socially embedded discourses on subjectivity.

Albert Newen and Kai Vogeley then extend the line from selfhood to the notion of a "first-person perspective." They do so by describing different cognitive abilities and introducing five levels of representation: nonconceptual representation, conceptual representation, sentential representation, meta-representation, and iterative meta-representation. These different types of representation are then put to work in an attempt to *operationalize* the concept of a first-person perspective—namely, as being involved in representational processes on different levels of complexity. Types of self-consciousness can now be developed into paradigms for the empirical investigation of neural correlates. Newen and Vogeley extensively discuss empirical studies that show converging evidence for a recruitment of medial cortical regions during mental operations of perspective-taking, even when operating on different degrees of complexity. Their discussion culminates in a speculative hypothesis: There exists a *neural signature for self-involvement* that is active independent of the degree of representational complexity to be performed.

Thomas Metzinger and Vittorio Gallese propose that having a consciously experienced first-person perspective is to have a certain type of representational content, namely an internal, phenomenal model of the intentionality-relation (PMIR). They investigate how such a model of transient subject–object relations could also be used for social cognition, empathy, and mind-reading by later developing into a model of subject–*subject*–relations. In particular, they propose that the F5 mirror system in the pre-motor cortex provided the minimal level of functional complexity out of which these high-level properties could unfold, because it can be demonstrated that it codes not object-presence, but the relational fact that the organism—or an external agent—is currently *directed* at an object component. However, the general question guiding Metzinger and Gallese is how the human brain could first develop a model of reality in which goals, actions, and selves are portrayed as distinct and individual elements, and what the specific contribution of the motor system to this achievement was. They close by discussing how the possession of such a neurally realized "action ontology" could later be successfully *shared* by groups of individuals, thereby enabling truly inter-subjective forms of interaction and opening the door into the social dimension.

(2) Self-recognition. The five articles in the second section explore the cognitive and brain bases of self-recognition. The first two contributions provide converging evidence for a right-hemispheric brain network that is dedicated to distinguish between self and others. The third contribution explores the role of proprioception, vision, and internal models for action identification. The last two contributions review behavioral studies, which suggest that the same cognitive system supports the processing of self-produced and other-produced actions.

Jean Decety and Thierry Chaminade provide a comprehensive overview of recent imaging studies addressing the neurophysiological substrate of self-recognition. Their review is guided by two assumptions. The first is that the same brain networks are involved in planning and imagining one's own actions, and in observing others' actions. Empirical support for this assumption comes from imaging studies showing that the same pre-motor and parietal areas are activated during action planning, action imagery, and action observation. Their second assumption is that a part of this brain network needs to be specialized in self-other distinction, because there are many social situations that require keeping self and other apart. Recent imaging studies suggest that the inferior parietal cortex in the right hemisphere serves this ability.

Dirk Leube, Günther Knoblich, Michael Erb, and Tilo Kircher report an imaging experiment that investigated the brain mechanisms of self-recognition. They devised a new task that aimed at eliciting the experience of an "anarchic hand" in healthy individuals. This task created an abrupt de-synchronization of the hand movements the person produced and the hand movements the person observed. In particular, individuals observed either their own hand or somebody else's hand move on, after they had stopped their actual movement. A right-hemispheric fronto-parietal network was selectively activated in the former condition. This network seems to play a central role in detecting mismatches between self-performed movements and their visual consequences.

Chloé Farrer, Nicolas Franck, Jacques Paillard, and Marc Jeannerod report two experiments that explored the role of proprioception, vision, and internal models in action recognition. In the first experiment participants either actively or passively moved a joystick. The visual feedback for the movement was spatially distorted. Conscious detection of the distortions was hardly impaired for passive movements as compared to active movements. This result suggests that action

recognition relies mainly on proprioception. The second experiment compared the performance of a de-afferented patient on the same task with the performance of healthy individuals. The patient's performance was clearly impaired. However, her detection rate still increased as the spatial deviation between performed and observed movements increased. This result suggests that action recognition does not fully rely on proprioception. Rather, it seems that internally generated predictions of action consequences can replace proprioceptive information.

Günther Knoblich and Rüdiger Flach address a further aspect of self-recognition, the ability to recognize one's own past actions. They assume that this ability reflects the workings of a common-coding system for action control and action perception. Perceiving one's own past actions leads to a higher activation in this system, because they are more similar to the codes in this system than others' actions. The higher similarity also increases the accuracy of the prediction of future action outcomes, because it provides a more informative input for action simulation mechanisms. A review of recent empirical evidence that addressed self-recognition of one's own past actions and the prediction of their future consequences provides support for this view.

Scott Jordan proposes an event-control account of the role of the self in human behavior. He argues that perception is inherently intentional. He reviews recent empirical evidence to show that perceptual and action-planning processes share representational resources. In addition, he reviews studies suggesting that the processes by which we attribute actions to "our-self" and the processes by which we attribute the actions of others to these others' selves are basically the same. Jordan concludes that it may be possible to map the concept "self" onto the regularities between perception and action referred to in the event-control model.

(3) Distortions of Self-Recognition. The three contributions in this section address distortions of the self in schizophrenic patients, especially those suffering from delusions of control. These patients often confuse self-produced and other-produced actions and sensations. The authors share the assumption that such delusional symptoms might reflect the failure of a self-monitoring system. The first two articles focus on action identification. The third article explores the relationship between self-monitoring and memory for actions.

Sarah-Jayne Blakemore starts from the assumption that forward models are crucial for the ability to distinguish between self-generated events and externally generated events. These models are part of the motor system and predict the sensory consequences of self-produced actions. Self-generated events can be distinguished from external events, because the predictions attenuate the corresponding sensory information. Accordingly, impaired forward models lead to a lack of attenuation for self-generated events, and thus, to an inability to identify them, as seen in schizophrenic patients with delusions of control. Blakemore then asks which brain systems might be dysfunctional in such patients. Her review of recent empirical evidence, including studies on patients with related neurological disorders and imaging studies on hypnotized individuals, suggests that an over-activity of the parietal cortex and the cerebellum during self-performed actions might be causing delusions of control.

Tilo Kircher and Dirk Leube stress the importance of more general conceptual frameworks for integrating empirical data from normal and clinical populations. They introduce such a framework, which distinguishes between different levels of consciousness and between different types of experiences. The second part of their article provides a more elaborated treatment of "self-agency," including a review of several related behavioral studies, imaging studies, and patient studies. They conclude that schizophrenic patients suffer from a failure in self-monitoring. This

conclusion is in line with the forward model account provided by Blakemore. However, Kircher and Leube suggest that a failure of self-monitoring might not uniquely characterize schizophrenic patients with delusions of control. Rather, schizophrenic patients with formal thought disorders might suffer from similar problems.

Elena Daprati, Daniele Nico, Nicolas Franck, and Angela Sirigu explore the role of self-monitoring and awareness of actions in memory encoding. They suggest that a failure in self-monitoring and the related reduction in action awareness in schizophrenic patients might affect memories that include the remembering person as the agent. Such memories are known to have a special strength in healthy individuals. The article reviews two experiments that tested whether this is also true for schizophrenic patients. The first experiment addressed source memory. Schizophrenic patients and controls were asked to indicate whether they had read an item aloud or silently during encoding. Schizophrenic patients were less accurate than controls. The second experiment addressed the enactment effect (better memory for actions that were performed during encoding). Other than healthy controls, schizophrenic patients did not exhibit this effect. Both results support the assumption that action awareness influences the consolidation of self-related memory traces.

(4) Volition, intention, and the self. The three contributions in the fourth section provide a collection of original experiments addressing the relationship between volition, intention, and the self. The first contribution explores the influence of thought instructions on the subjective experience of intentionality. The other two contributions used Libet's method to explore the perceived timing of self-produced and other-produced intentional actions and their effects, and the perceived timing of involuntary movements.

Dan Wegner and James Erskine raise the question of whether one can voluntarily behave involuntarily, or intend not to intend. This possibility is suggested by their theory of "apparent mental causation" which is inspired by Hume. According to this theory, "the experience we have of causing our own actions arises whenever we draw a causal inference linking our thought to our action." Thus, mental control of the availability of thoughts might "potentially undermine the person's experience of voluntariness during subsequent action." To test this possibility, Wegner and Erskine conducted an experiment in which they instructed the participants to either suppress thinking about their intention, to concentrate on their intention, or to monitor their thoughts, before they carried out an action. The results showed that the instruction to concentrate or monitor increased the subjectively experienced intentionality, whereas the instruction to suppress tended to reduce the subjectively experienced intentionality.

Patrick Haggard and Sam Clark explore the relationship between the brain processes, which underlie intentional action, and the phenomenal experience intentional action induces. In their experiment they used transcranial magnetic stimulation to distinguish between two alternative models of this relationship: A Humean model, according to which a retrospective inference induces the phenomenal ownership of the executed action, and a constructive model, according to which the phenomenal ownership of the executed intentional action originates from the predictive link established between intentions and their physical effects in the world. Haggard and Clark conclude that the intentional binding effects observed in their experiment provide evidence for the constructive model.

Andreas Wohlschläger, Kai Engbert, and Patrick Haggard report two experiments addressing the role of intentionality and proprioception for the awareness of self-produced and other-

produced actions. In these experiments, participants judged the time of a key-press on a Libet clock. The movements were either intentional or unintentional, and they were or were not accompanied by proprioceptive feedback. The results showed that the availability of proprioceptive information affected the perceived timing of unintentional movements, but did not affect the perceived timing of intentional actions. This implies that the perceived timing of self-produced and other-produced intentional actions did not differ. The authors conclude that when an action produces an intended effect, "the differential activation of the proprioceptive system seems to be overridden by mechanisms that attribute intention."

(5) Self-development. The fifth set of contributions tries to shed some light on the development of self and agency in the first years of life. Behavioral observations in infants and young children revealed that the self is not innate, but gradually develops within the first years of life. Different sources of self-knowledge develop at different points in time. Many of these sources are tightly connected to the developing sense of agency. By actively exploring the environment, infants and young children learn to recognize themselves as agents among other agents, and as a self among other selves.

Philippe Rochat reviews findings from developmental research and identifies five levels of self-awareness that unfold from birth to five years of age. The general idea is that prior to the expression of explicit self-awareness, which becomes evident in the second year by self-recognition in a mirror, infants manifest an implicit sense of themselves, that is, they perceive their own body as a differentiated entity among other environmental entities. Rochat conceptualizes self-awareness as a dynamic process that expands from the perception of the body in action to the evaluative sense of self as perceived by others.

Birgit Elsner and Gisa Aschersleben assume that infants' ability to learn about the consequences of actions is a prerequisite for developing a self. They report an experiment in which they investigated whether infants learn about the effects of other persons' actions like they do for their own actions. Children in the age range of 9–18 months either observed or did not observe an adult model, before they explored a novel object themselves. By 12 months, but not earlier, the infants benefited from observing the model's actions and their effects. By 15 months, infants expected their own actions to produce the same effects as the model's actions. Elsner and Aschersleben conclude that the emergence of explicit self-awareness in the second year of life is accompanied by a new way of understanding others' actions.

Ildikó Király, Bianca Jovanovic, Wolfgang Prinz, Gisa Aschersleben, and György Gergely examine how infants in the first year of life understand the actions of other persons. They claim that certain features of observed actions, like the equifinal variation of action and a salient action effect, lead young infants to interpret actions as goal-directed. A study with 6- and 9-month-old infants provides evidence for their assumption. To account for these findings, they propose a theory according to which infants are equipped with an action interpretation system, the teleological stance. This system is specialized in representing goal-directed actions. Thus, very early in life, infants expect that the basic function of actions is to bring about some particular change of state in the world, and that agents will use the most efficient means available to achieve such changes.

Beate Sodian, Christian Hülsken, and Claudia Thoermer address self-development in school-age children. They claim that theory-of-mind research contributes to the understanding of the relation of self and action (1) by exploring the relation of the development of self-knowledge and

of knowledge of others' minds and (2) by investigating the relation between theory-of-mind development and the development of action control. They illustrate these claims with a review of some recent empirical evidence, including a study on theory-of-mind in children with deficient action control (ADHD-diagnosed children). Beate Sodian and her colleagues suggest that theory of mind leads to improved action control which in turn supports the ability to represent mental states on-line.

We hope that the collection of articles in this special issue will inspire those interested in exploring the cognitive and neural processes that define our selves. The majority of articles collected in this issue originate from the talks and discussions given at the symposium "Self and Action" (October 2002, Ohlstadt, Germany) that was made possible by the generous support of the Max Planck Institute for Psychological Research, Munich. Further articles and commentaries were later invited to complete the picture. We would like to express our gratitude to the editors of Consciousness and Cognition for providing us with this prestigious platform, and for their support during the preparation of this special issue. William Banks was of enormous help at all stages of this process, and Bernhard Baars and Antti Revonsuo provided very helpful input on earlier versions of the manuscripts.

References

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.