



ELSEVIER

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

I move, therefore I am: A new theoretical framework to investigate agency and ownership

Matthis Synofzik^{a,*}, Gottfried Vosgerau^b, Albert Newen^b

^a Centre for Neurology, Hertie-Institute for Clinical Brain Research, University of Tübingen, Hoppe-Seyler Strasse 3, 72076 Tübingen, Germany

^b Department of Philosophy, Ruhr-University Bochum, Germany

ARTICLE INFO

Article history:

Received 28 February 2008

Available online 14 April 2008

Keywords:

Self-representation

Agency

Ownership

Alien hand

Anosognosia

ABSTRACT

The neurocognitive structure of the acting self has recently been widely studied, yet is still perplexing and remains an often confounded issue in cognitive neuroscience, psychopathology and philosophy. We provide a new systematic account of two of its main features, the sense of agency and the sense of ownership, demonstrating that although both features appear as phenomenally uniform, they each in fact are complex crossmodal phenomena of largely heterogeneous functional and (self-)representational levels. These levels can be arranged within a gradually evolving, onto- and phylogenetically plausible framework which proceeds from basic non-conceptual sensorimotor processes to more complex conceptual and meta-representational processes of agency and ownership, respectively. In particular, three fundamental levels of agency and ownership processing have to be distinguished: The level of feeling, thinking and social interaction. This naturalistic account will not only allow to “ground the self in action”, but also provide an empirically testable taxonomy for cognitive neuroscience and a new tool for disentangling agency and ownership disturbances in psychopathology (e.g. alien hand, anarchic hand, anosognosia for one’s own hemiparesis).

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Throughout our everyday life we continuously perform actions, often accompanied with the phenomenal experience of “mineness” of our own actions.¹ What does it mean, however, to experience an action as *my* action? The following work aims at developing a comprehensive naturalistic account of the experience of “mineness” of an action (which, as will be shown, actually requires two different accounts: One account of the experience of *agency* and one of the experience of *body ownership*) that can be easily integrated into a naturalistic theory of the self and of self-consciousness. This account not only serves as a purely conceptual framework, but—as will be demonstrated in the last section—also provides an experimentally testable taxonomy for neuroscience and cognitive science and an explorative tool for investigating neuropsychological disturbances of agency and ownership. We will argue that, in particular, one has to differentiate (i) an individual-orientated cognitive dimension of agency and ownership from a socio-normative dimension of agency and ownership. Moreover, within the individual-orientated cognitive dimension, one has to differentiate (ii) between a feeling and a judgment of agency as well as between a feeling and judgment of ownership. As an exemplary use, it will be shown that this distinction is very fruitful to analyze and re-describe psychopathological case studies.

* Corresponding author.

E-mail address: matthis.synofzik@uni-tuebingen.de (M. Synofzik).

¹ This experience occurs mostly in an implicit, unconscious way; only under special conditions of voluntary attention do we become explicitly aware of what we are attempting to do and of ourselves as the initiator of these actions (Legrand, 2007; Synofzik, Vosgerau, & Newen, 2007).

2. Features of self-consciousness: Who are we?

Self-consciousness typically comprises certain characteristic features of our experiences:

1. The unity of our experiences,
2. The perspectivity of our experiences,
3. The sense of ownership of body parts,
4. The sense of agency of actions,
5. The sense of authorship of thoughts, and
6. The transtemporal integration into autobiographical knowledge (Bermúdez, 1998; Damasio, 2000; Metzinger, 2003; Newen & Vogeley, 2003).

All these phenomena can be characterized on both an individual cognitive level and a social and normative level (as was demonstrated e.g. with respect to emotions by Zinck and Newen, 2007). In this paper, we will describe both dimensions for agency and ownership.

It was traditionally assumed that there has to be a distinct, non-reducible entity called *self* which would underlie these core features, functioning e.g. as the realizer of the unity of our experiences and of transtemporal unity, as the center of our perspective or as the bearer of the sense of mineness of actions and thoughts. Recently, an alternative approach that tries to account for the phenomenology of self-consciousness has received widespread attention: The claim that there is a phenomenally quite unitary and central *self-model* (at least in non-pathological circumstances), but no actual *self*. According to this approach, the supposition of a self has to be denied since it is always thought of as non-natural. It seems, however, as if both traditional positions, like those put forward by Descartes and Kant, as well as recent accounts, such as Thomas Metzinger's *Being no one* (2003), have overlooked an important alternative by neglecting the difference between the representing *self* and the represented *self-model*²: While the *self* is—at least from a naturalistic view—best understood as the objective, biological-cognitive representational system with special characteristic self-representational capacities, the *self-model* is the subjective representation of oneself constructed³ by one's self-representational system based on one's perceptions, emotions, beliefs, etc. (Newen & Vogeley, 2007). Under functional and representational levels of description, self-representations (e.g. a self-model) are not build up by a homogenous monolithic entity—as it appears phenomenally—, but by different modules operating on different levels of representational and functional complexity.

Within a naturalist account of human self-consciousness (defined as the ability to represent one's own bodily and mental states as one's own states), a recent systematic analysis has been able to identify different levels of self-representation characterized by the increasingly complex representational structure and the increasingly complex representation formation processes (Newen & Vogeley, 2003). On these grounds, we have to distinguish at least three different levels: A non-conceptual level, at which states are represented in a non-analyzable way (including an only implicit self-representation), a conceptual level,⁴ at which properties can be systematically ascribed to different objects (including the explicitly represented self), and a meta-representational level, where mental representations can themselves be represented (and attributed to the self or to others).

The advantage of this framework is that the complex phenomenon of self-consciousness is now characterized by specific gradually proceeding systematic competencies that can be distinguished by (i) their formal representational structure, (ii) their developmental acquisition in childhood development, (iii) their differential break-down in certain neurological or psychopathological circumstances, (iv) their behavioral correlate in psychological paradigms and (v) their neural correlate in neuroimaging or lesion studies.

Not only self-consciousness as a whole, but also each of its core features resorts to analogous systematic competencies that can be distinguished by the same (or at least similar) individuation criteria which also evolve in gradual steps. This will be shown in this paper by example of the experience of agency and of ownership. In other words: Although each of the core features of self-consciousness appears as a phenomenally unitary and central processing module, they in fact consist of rather complex crossmodal phenomena of largely heterogeneous functional and (self-)representational levels, where the most basic levels consist in sensorimotor processes, while the most elaborated levels comprise conceptual and meta-representational processes. By identifying and characterizing each of the systematically relevant levels, we obtain a promising method to “naturalize” complex mental phenomena, a challenge which formerly seemed to be insurmountable.

² While traditional positions presuppose only a self, portraying it as a non-reducible entity (a Cartesian *res cogitans* or a Kantian transcendental 'I'), Metzinger assumes only a self-model, excluding any additional use of self. In a naturalist view, however, it seems necessary to adopt both, the self and the self-model, in order to account for self-related pathologies.

³ Although the self-representation is “constructed” by the self-representing system, this does not imply that it is a mere fantasy without any represented entity in the real world (as apparently connoted by Metzinger's usage of the term). Rather, this construction reflects facts of a genuine object of the world: the self. Rather than “being no one” (Metzinger, 2003) we are *self-representing cognitive systems*.

⁴ Newen and Vogeley (2003) make a distinction between conceptual and propositional representations: For the latter also sentential connectives have to be represented. This distinction will not play a role here such that our usage of “conceptual” comprises propositional representations.

Table 1

Levels of the sense of agency and the sense of ownership

	Sense of agency	Sense of ownership
Non-conceptual and momentary representation		Sensory registration of action–effect-couplings
Non-conceptual and stable representation	Feeling of agency	Feeling of ownership
Conceptual representation	Judgment of agency	Judgment of ownership
Meta-representation	Ascription of moral responsibility	Meta-representation of ownership

3. What are we experiencing: Agency and ownership

In our everyday experience of our voluntary actions we have a certain phenomenal experience of “mineness” of these actions. For example, when reaching for a cup, I know this to be my action. This experience, however, involves at least two fundamentally different aspects of minimal action-related self-awareness, which—in this case—coincide and are phenomenally indistinguishable: The sense of agency (SoA), i.e. the registration that I am the initiator of my reaching for a cup (rather than somebody or something else),⁵ and the sense of body ownership (SoO), i.e. the registration that my arm which is reaching for a cup belongs to me (rather than belonging to somebody or something else). In more general terms, ownership is the registration that parts of one’s own body belong to oneself, regardless of whether it is moving or not and whether this movement is voluntary or involuntary.⁶ In case of passive movements, however, SoA and SoO dissociate. If someone else moves my arm towards a cup, I have no longer a SoA (i.e. I no longer experience myself as being the initiator of the action), but I still retain a SoO: The arm that is moving towards the cup is still experienced as belonging to me, i.e. as *my* arm (Gallagher, 2000).

As will be shown in the remainder of the paper, these two senses—which are often confounded both in theoretical and experimental work (for further analysis, see Synofzik, Vosgerau, and Newen, 2007)—are *analog* in their formal representational structure in that each of them can be framed in a comprehensive multi-level framework which gradually increases in representational and functional complexity and proceeds from basic non-conceptual sensorimotor representations to full-blown conceptual representations of agency and ownership, respectively. Each level of the SoA and the SoO, respectively, will be individuated by means of the following individuation criteria: (i) Its form of representation, (ii) the cognitive capacity it makes use of, (iii) the underlying neurocognitive process and (iv) its differential break-down in certain neurological or psychopathological circumstances. The different levels interact with each other by both bottom-up and top-down-processes. The SoA and the SoO are, however, clearly *distinguishable* not only on conceptual, but also on phenomenal, neurocognitive and psychopathological grounds. By means of the aforementioned criteria of individuation one can identify at least four systematic levels of the SoA and the SoO, respectively (see Table 1).

4. Preconditions of agency and ownership: Sensory registration of action–effect-couplings

4.1. General characteristics

In order to experience oneself as the initiator of one’s own actions, it is a very basic prerequisite that a system learns to systematically register a certain sensory event as the typical effect of its own actions. This can be realized on the basis of two different kinds of non-conceptual representations: On a very basic level, this action–effect-registration results in a situated and momentary representation. What is learned on this level is to represent a certain sensory event as the *effect* of a certain type of action (which is *de facto* my action, since there is no systematic covariation of other events), thereby establishing a basic momentary self–environment distinction which includes representing the effect as produced by me and by my body parts (for a more detailed explanation of the role of action–effect-couplings for building up a self–environment distinction see Vosgerau & Newen, 2007).

The self-relation⁷ is, although *de facto* present, not yet represented. So, what we accomplish at this basic level of sensory covariation is a momentary self–environment distinction which nevertheless allows a basic representation of my action. On the phenomenal level, we already have a momentary sense of initiating an action, although we cannot yet speak of a sense of agency (it is not represented as *my* action). This can be nicely illustrated by looking at persons suffering from a loss of proprioception (patient GL, see Farrer, Franck, Paillard, & Jeannerod, 2003; Fournieret, Paillard, Lamarre, Cole, &

⁵ For a more detailed explanation of this definition of the SoA and its difference as compared to other currently often used definitions (e.g. the definitions proposed by Gallagher, 2000; Frith, 2005 or Tsakiris and Haggard, 2005a, see Synofzik et al., 2007).

⁶ A distinction between a sense of agency and a sense of ownership was first suggested by Gallagher (2000). The definitions and neurocognitive characterizations we use, however, differ in many important aspects from Gallagher’s definitions and characterizations (see Synofzik et al., 2007). Nowadays, this distinction has already become rather common in the current literature (see, for example, de Vignemont and Fournieret (2004) or Tsakiris and Haggard (2005a)).

⁷ The self-relation is always understood as an essential self-relation, i.e. a *de se* and not simply a *de re* relation.

Jeannerod, 2002). If such a person cannot see her hand but starts to move it forward, she develops a momentary sense of initiating the action, but since there is no ongoing control of the hand movement there is no stable feeling of agency. On this level, the sensory action effect representation can be best described as a non-conceptual and momentary representation “This is an initiation of action”.

This most basic level functions as a registration of action–effect-couplings and is a necessary prerequisite for both the SoA and the SoO. It is a non-conceptual, momentary representation of oneself being the agent which can be used by the cognitive system for *multiple* purposes. For example, action–effect-registration is useful for building up basic goal-representations, for starting basic action planning and especially for dealing more efficiently with self-induced sensory stimulation. Moreover, the information conveyed by the action–effect-registration can also be *recalibrated* in different ways, depending on what the task or context requirements are. For instance, when action–effect-registration functions as part of the integration of self-induced sensory stimulation, it can serve either to *cancel* the self-induced effect (thus preserving higher attention resources for world-induced sensory stimulation in sensory-overloaded contexts), to *attenuate* the self-induced effect (thus still allowing perceptual awareness of it in contexts when further cognitive control is required) or to *highlight* the self-induced effect (thus allowing to disambiguate one’s own action effects in ambiguous, complex action contexts). Thus, action–effect-registration serves many heterogeneous pragmatic, executive functions.

Moreover, the basic registration of action–effect-couplings is also a necessary prerequisite for the SoO as well as the SoA: In order to experience a certain body part as part of one’s own body it is a very basic prerequisite that a system learns to systematically register certain sensory inputs as the sensory inputs that arise from *one’s own* body (rather than from someone else’s body or from the world). It allows to construct a momentary (non-stable) representation of the mineness of the body parts. Only in the course of acting, we can learn to reliably distinguish the sensory information which belongs to ourselves and that which does not, since only in this case we can register systematic covariations between sensory inputs and our actions and thus distinguish it from sensory input just coincidentally affecting us from the surroundings. Nevertheless, on this level, the distinction between ownership and agency is not yet made by the system itself (i.e. it is not represented); rather, both aspects of the mineness of the action are present in the momentary non-conceptual representation of me being the agent.

4.2. Neurocognitive mechanism

The action–effect-registration presupposes only very parsimonious requirements to the action control capacities: First, the system must start to move somehow, i.e. it just has to perform spontaneous, non-intended, non-goal-directed movements. Children are known to execute such spontaneous movements even long before birth. Thus, contrary to all the following stages, no special capacity of action control is actually needed so far (not even an action prediction—as needed, e.g. for the instantiation of a comparator mechanism; see below). Second, the system must have the ability to detect and store systematic covariations—an ability inherent to neural network processing from its earliest stages (Hebb, 1949). No other cognitive capacity and no other neurocognitive mechanism is needed. Thus, the explanation of the registration of action–effect-coupling relies on unproblematic presuppositions.

The empirical plausibility of such a registration capacity and its distinction from the next, more complex level is well-supported by several studies from developmental psychology as summarized, e.g. by the developmental two-stage model of intentional agency from Hommel and colleagues (for review see Hommel & Elsner, *in press*): On this first level, non goal-orientated reflexive movements are performed and the simultaneously occurring sensory events are simply explored. If certain movements and sensory events occur often together and the movements are performed over and over again, simple learning mechanisms cause an action–effect association learning (Elsner & Hommel, 2001). For example, 2-month-old infants can modulate their oral activity (sucking on a dummy pacifier) to increase the optical clarity of a film presented to them or to control the pitch variation of a sound (Rochat & Striano, 1999). Furthermore, 2–6 month-old infants can, on a very basic level, monitor and control their kicking and reaching behavior on the basis of these action–effect contingencies, as illustrated, for example, by the finding that they pedal longer if this leads to a systematic consequent movement of a mobile (Hommel & Elsner, *in press*; this example well illustrates the aforementioned thesis that action–effect-registrations on this level are not yet necessarily used for installing a phenomenal experience of self-agency, but rather serve many pragmatic executive functions). Only when action–effect-associations are reliably learned and ready for (inverted) use for the motor system, intentional agency can develop: While the associations are *acquired* through executing movements and exploring their sensory effects (motor activity → sensory system), they are a necessary *prerequisite* for the goal-oriented tuning of intended actions (imagined sensory state → motor activity) (Hommel & Elsner, *in press*). Since the latter inverse representations and, more generally, *internal* representations of actions and action effects are not yet systematically established and the self-relation is not represented, the registration still depends on the immediate presence of the action and action effect. Hence, the underlying action representation lacks transtemporal stability. Correspondingly, the ability to act in a truly goal-oriented fashion and to anticipate action consequences of ones own actions—both hallmarks of the next level of agency, the feeling of agency—emerge much later, namely from 9 months of age onwards, than the capacity to detect sensorimotor contingencies (Hommel & Elsner, *in press*; Tomasello, 1999).

5. Dimensions of agency

5.1. Feeling of agency

5.1.1. General characteristics

If the action–effect–registration is systematically used by the cognitive system to establish a stable representation of an action as self-produced, then the non-conceptual feeling of being the agent is produced (feeling of agency, FoA). What is learned on this level is to establish a stable perception-based representation of the effect of one's own action as the effect of one's own action. The self-relation is now represented, although still in a non-conceptual (implicit) way. This agency-representation is in some way abstracted from the action such that the feeling of being the agent still is present even if the action is already over. Such a feeling can be persistently accessed and attended to although it can be established independently from any conceptual, a fortiori linguistic representations. Based on the weighting process of action-related motor and sensory cues on this level (see below), one can now establish a stable representation of the “mineness” of the action–effect such that it becomes accessible for other non-conceptual cognitive sub-systems. An action is, however, merely classified as self-caused or not self-caused—it is not *attributed* to the self. The self is only represented non-conceptually in the FoA because at this level the representation is still not composed out of different parts for the self and for the property of being the cause. Rather, the content “This is my action” is rather present or absent as a non-analyzable whole. Nevertheless, the implicit self-representation being established on this level is a stable and abstract self-representation as opposed to the basic self-environment representation relying on action–effect-couplings.

Importantly, an external attribution is still not possible at this level. Therefore, no such thing as a feeling that, say, Chris has caused my action, can occur. Rather, in case of an inappropriate action effect, the action feels peculiar or strange, while in case of an appropriate action effect, I feel a harmonious ongoing flow of my action sequence. Perceptual agency representations are not compositional and have no object-property structure, and are therefore non-conceptual (Newen & Bartels, 2007; Vosgerau, 2007). They do not allow for explicit representations of the self and hence not for attributions.

5.1.2. Neurocognitive mechanism

The FoA is produced by a gradual and highly plastic subpersonal weighting process of different action-related perceptual and motor cues, which are partly afferent (visual feedback, proprioception) and partly efferent (action prediction, body schema) (see Synofzik et al., 2007, and Fig. 2). In case of *congruency* between these indicators (e.g. a match between internal prediction and visual feedback), the experience of action is withheld from further processing and we experience self-agency by a rather diffuse feeling of a coherent, harmonious ongoing flow of action processing. In case of *incongruency* between these indicators (e.g. a mismatch between proprioception, motor intention and visual feedback), we experience an action as strange, peculiar and not fully done by me. This experience might—inter alia—result from a mismatch between efferent and afferent information.

One of the most important authorship cues is provided by a presupposed internal comparator mechanism which compares a prediction about the sensory consequences of our action (computed on the basis of an efference copy issued as a “corollary discharge” of the motor command of the respective action) with the actual sensory feedback of the action (see Fig. 1). A *congruency* between predicted and actual state may be used by the cognitive system to register a sensory event as caused by oneself, while an *incongruency* may lead to the registration of a sensory event as externally caused (Feinberg, 1978; Frith, 1992); for a recent version and modification see (Frith, 2005; Frith, Blakemore, & Wolpert, 2000).

The comparator functions as one of the most important sensorimotor cues for the FoA. The acquisition and adaptation of the internal predictions (which are entered into the comparison process together with the sensory afferences) can be described as the result of an “internalization” process of one's own causal relationships to the world: The representation of stable systematic sensorimotor contingencies between a certain action and its corresponding effect, which is learned in the preceding stage of registration of action–effect-couplings, can be used to build up an internal prediction of an action effect that is issued whenever an action is initiated (Synofzik, Thier, & Lindner, 2006; Vosgerau & Newen, 2007). If the action effect matches the prediction, the prediction is preserved; if the action effect differs slightly over a constant sequence of actions, the prediction is recalibrated accordingly (neurophysiological evidence for this mechanism was given for electric fish (Bell, 2001), crickets (Poulet & Hedwig, 2002) and also for humans (Haarmeier, Bunjes, Lindner, Berret, & Thier, 2001; Lindner, Haarmeier, Erb, Grodd, & Thier, 2006; Synofzik, Thier, & Lindner, 2006)). The more systematic the contingencies between action and action effect, the more stable and reliable the internal predictions about the upcoming sensory effects of one's own actions and the better a system's capacity to differentiate between self-produced and non-self-produced actions. Thus, the cognitive system is now no longer dependent on the presence of actual sensory feedback of the action (as for the registration of action–effect-couplings), but can internally represent the action consequences by internal predictions which thus serve as “internal feedback signals” (Ito, 2005).

Yet there are many cases where the comparator output is neither a sufficient nor a necessary condition for the FoA (Synofzik et al., 2007). Action-related perceptual and motor information—such as efference copies, sensory feedback modalities (e.g. proprioception (Balslev, Cole, & Miall, 2007) or vision (Synofzik, Thier, & Lindner, 2006)), and their comparison—is rather entered into a multifactorial weighting process bringing together different sense modalities (e.g. intermodal congruencies), efferent–afferent congruencies, and different levels of cognitive processing that indicate agency of actions for a person (Syn-

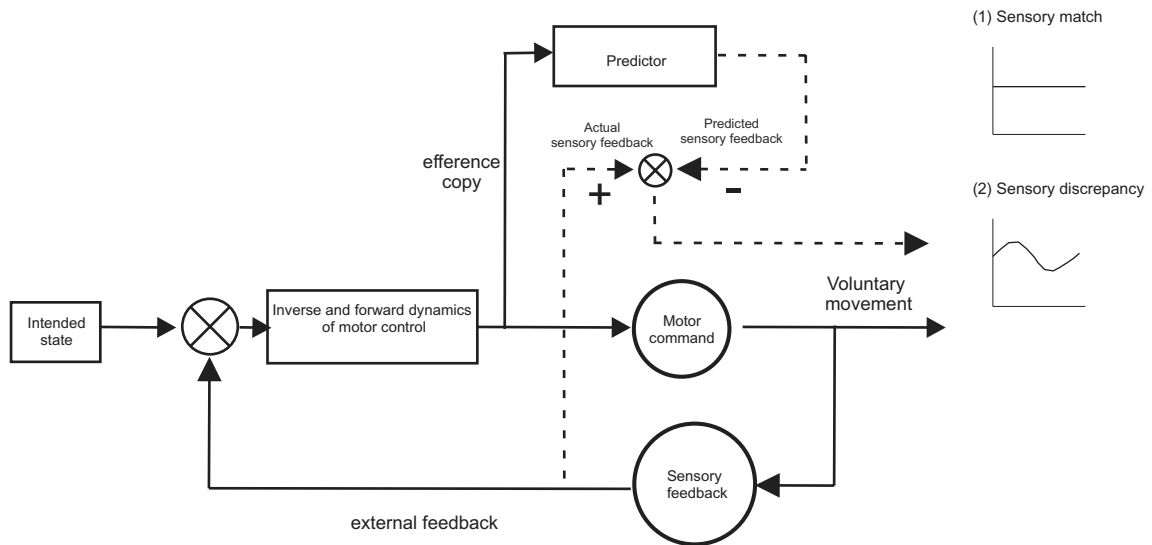


Fig. 1. The “comparator mechanism” as an important authorship indicator for the feeling of agency. The feeling of agency is—inter alia—informed by an internal comparator mechanism which compares predicted and actual sensory feedback: If both states match, the sensory event is attributed to one’s own agency. If sensory feedback is discrepant with the sensory prediction, an external attribution of the causation of the sensory event occurs (modified after Frith et al., 2000 and Wolpert & Flanagan, 2001).

ofzik et al., 2007). The differential contribution and the weighting of the corresponding authorship cues vary in different situations and in different persons (for more details, see Synofzik et al., 2007).

5.2. Judgment of agency

5.2.1. General characteristics

If the non-conceptual FoA is further processed by the cognitive system by additionally involving conceptual capacities and belief stances, then a conceptual, interpretative judgment of being the agent (judgment of agency; JoA) is produced. What is learned on this level is to *conceptually* represent the effect of one’s own action as the effect of one’s own action. Based on the belief formation process of weighting different cognitive, *non-sensorimotor* indicators (see below), one can now not only perceptually represent the difference between “my action” and “not my action”, but propositionally and compositionally represent one’s own actions as one’s own or as the one of somebody else, i.e. the self is now explicitly represented as well as the property of being the cause. Therefore, on this level, attributive judgments (e.g. “my action” vs. “the action caused by Chris” or “the action initiated by the computer”) are possible. For each action, an interpretation mechanism can search for the best explanation of the underlying sensorimotor perception, resulting in a specific belief formation about the origins of the changes in perception: I can either believe that I am the author of the action or I can assume that someone or something other than me is causing this effect.

By this characterization, the JoA differs from the FoA in the following three, partly interrelated respects:

- (1) The JoA has an *object-property structure*, i.e. there are parts of the conceptual self-representation that represent the system itself and parts that represent certain properties. Thus, whereas the FoA involved a non-conceptual (implicit) self-representation, the JoA involves a *propositional I-thought* (with an explicit self-representation).
- (2) Since conceptual agency representations are—contrary to the FoA—formed by *inferential processes*, they are influenced by other conceptual representations like *background beliefs*. Consider, for example, you are the only person in a room, in which something happens. Although you might not have any perceptual representation pointing in the direction that you caused this event (e.g. no certain congruency between motor intention and sensory event), you can come to conclude that it must have been caused by you because of your background beliefs that (a) nothing happens without a cause and (b) you are the only possible cause in the relevant spatiotemporal region (de Vignemont & Fourneret, 2004). Thus, how the belief formation is performed depends on our way of *rationalizing*, i.e. our way of giving a (more or less) plausible explanation for our experiences.
- (3) While the FoA builds on the cognitive capacity to *classify sensory events* as one’s own or not one’s own, the JoA demands the capacity to *conceptually categorize causal forces* in the interaction with the world. The conceptual categorization has to meet different constraints (Newen & Bartels, 2007): The system has to have the ability (a) to attribute the same causal force to different objects, (b) to attribute different causal forces to one and the same object, (c) to perform the attribution stimulus-independently, and (d) to distinguish enough related properties such that a network of different causal dimensions is established (i.e. it has to be able to differentiate between agency and other related properties, e.g. ownership).

5.2.2. Neurocognitive mechanism

The JoA is formed by a rationalization process that normally has a FoA as input. Thus, the JoA normally starts with the output of the weighting of *sensorimotor authorship indicators* (e.g. the comparator output) and then proceeds with ad hoc theorizing and belief formation about oneself, which is produced by a weighting of *cognitive authorship indicators* (e.g. contextual cues, belief states, etc.). But the JoA does not necessarily presuppose a FoA. The rationalization process has multiple inputs such that it works even if the FoA input signal is missing: Analogous to the sensorimotor weighting, also the weighting of the various cognitive authorship cues varies in different situations and in different persons. For example, in the aforementioned example of a person being alone in a room, the cognitive system draws mainly on *situational, impersonal background beliefs* to decide whether an action effect is caused by oneself or not: The background beliefs about causation and critical spatiotemporal patterns that are used in this situation to achieve a JoA are so general, that they do not depend on the person's individual biography and psychosocial context. For deciding, however, *to whom* agency is externally attributed—if it occurs—(e.g. to decide whether an action was caused by my friend Paul, by the computer, by the experimenter, by God, etc.), the cognitive system must often draw on more specific *personal background beliefs* (which are, for example, stored in form of a narrative) and *social cues*.

Next to these examples, additional empirical evidence for a cognitive weighting process is given in an indirect way by the observation of the break-down of this process, e.g. in the case of agency delusions. Here, neuropsychological anomalies in perceptual or affective processing are insufficient to explain the intrinsic nature of delusions, namely the fact that (i) an alien experience is not accepted as a strange experience, that (ii) a delusional agency hypothesis about the perceived fact is formed and that (iii) this is maintained despite different stored encyclopedic knowledge about the own behavior and despite the testimony of others (cf. Davies, Coltheart, Langdon, & Breen, 2001). These features can only be explained by a misbalanced integration of different cognitive cues, in particular background beliefs, contextual cues and action intentions.

Like the weighting of sensorimotor authorship indicators, also the weighting of cognitive authorship indicators usually remains unconscious and is performed on a subpersonal level. Thus, although the belief formation involves reasoning processes or narrative elements in order to proceed from a certain perceptually accessible premise (the result of the sensorimotor weighting) or set of personal belief states to the conclusion that oneself or somebody else is the initiator of the action, this does not necessarily have to be performed in a conscious way. Nevertheless, such a conscious evaluation is possible and likely to occur when the discrepancy between several factors is significant.

On the level of JoA, a system is able to represent its action consequences and action effects as stable in space and time by encoding them in form of propositional knowledge. These stable internal representations increase the system's degree of independency from direct (reafferent or exafferent) sensory input in that they are not just limited to predicting the very momentarily upcoming sensory consequences, but extend beyond direct sensorimotor consequences to more abstract action effects and beyond very immediate action consequences to spatially and temporally more distant action effects. This allows building up an increasingly abstract goal representation as well as explicit and systematic judgments about causal action and world relations. Thus, whereas the system was first capable of forming perceptually encoded internal representations of one's own direct sensory action effects, allowing the system to become independent from the actual presence of the action effect (level of action control corresponding to the FoA), it can now anticipate more abstract action effects and can furthermore do so without coupling them to a specific motor pattern or a specific action–control loop. In other words: Self-agency can now be represented not only independently from the actual *action–effect*, but also independently from the *actual action*.

5.3. Ascription of moral responsibility

5.3.1. General characteristics

If a cognitive system cannot only form agency beliefs and judgments, i.e. attribute a certain action to a certain author (in the sense of being the cause), but also reliably register the mental state preceding the action—namely, the intentional action planning—, it is capable of ascribing moral responsibility. The criteria for the ascription of moral responsibility (AoR) can be defined as follows: An agency system A (as a minimally complex behavioral system that is able to produce a behavior B) is morally responsible for its behavior B (with the consequences C) if and only if:

- (I) A has an internal action plan including a representation of the behavior B (and some consequences C),
- (II) A has sufficient insight into the normally possible consequences C of B,
- (III) there are no contextual barriers which prevent A from doing B and thereby realizing the normally expectable consequences C,
- (IV) A's doing B is embedded in a normative system, i.e. the doing is evaluated according to a normative rule as acceptable or non-acceptable.

Taken together, these conditions mark the specific presuppositions required for moral responsibility: Conditions 1–3 are necessary and sufficient for causal responsibility of a subject as an agent.⁸ Condition 4 is required to specifically mark the nor-

⁸ These three conditions go beyond being the mere cause, as characterized at the level of JoA where the plans of the acting system are irrelevant: I, as a living thing, am the cause of the consequences of my reflexes, for example, but I am not causally responsible for them as the agent of the action.

mative dimension of moral responsibility. The conditions shall be shortly elaborated: Criterion (I) allows us to separate actions from mere behavior: Actions are based on an internal action plan (which may be unconscious at the time of acting) including the consequences of the behavior (thereby excluding mere automatic reflex-like behavior or behavior caused by someone/something else). Criterion (II) demands that the subject knows sufficiently what consequences C are possible under normal circumstances. This excludes several kinds of behavior as reasonable candidates for responsibility, e.g. if the subject's ideas about the consequences of its behavior are so mistaken that they have almost nothing to do with the actual consequences—despite the fact that they might have been based on an internal action plan. Condition (III) grasps our strong intuition that I am not responsible for my behavior if there are special circumstances which prevent me from realizing the behavior according to my existing action plan (notwithstanding the fact that this plan might be appropriate to foresee the consequences which usually follow the action). Condition (IV) reflects our notion that responsibility essentially includes that one can be either blamed or honored for one's behavior and this is only possible if we have standard expectations according to norms in a group. Knowledge of social norms is closely connected with meta-representations of the form “I know that I should do A1 and should not do A2. Otherwise mum will be angry and disappointed.” Hence, normative expectations presuppose an understanding of other people's expectations about my actions: “She wants/expects me to do A1 and not A2.” In contrast, for a judgment of agency (JoA) no *meta-representation*, i.e. no thought about particular mental representations of either myself or of others, was necessarily required. All that is needed for a JoA is an attribution of an action to a certain agency system, which might, for example, be expressed in the linguistic form “He caused this action” or “This is his action” (him being the mere cause). Moreover, while the AoR resorts to a *normative standard* with an associated evaluation as acceptable or not acceptable, the JoA only involves a notion of causation allowing the subject to differentiate between oneself or someone else as being the cause of action effects.

The normative expectations which underlie the AoR are internalized and individually adopted normative rules that offer a standard of evaluation of the actions of a person. Since a normative rule is a rule determined by a culturally mediated social interaction, responsibility is an essentially culture-dependent phenomenon: While suicide bombing is morally acceptable in some radical ideologies, it is completely unacceptable in most cultural contexts (e.g. the Western European culture).

The meta-representational dimension of social interaction drawing on social standards and normative judgments has to be distinguished from a person's individual cognitive dimension of feeling and judgment, in that feeling (FoA) and judgment (JoA) are both ontogenetically and phylogenetically more basic features. They are ontogenetically more basic in that they do not require meta-representations and thus no construction of a model of one's own or someone else's mind (see next paragraph)—a capacity which is not learned until approximately the third year of life. They are phylogenetically more basic in that they do not require the construction of socio-cultural normative rules—a capacity which presupposes a special kind of competence that is developed in human social interaction, but not, for example, in ape interaction (“cultural intelligence hypothesis”). This fact was shown by recent studies in developmental psychology revealing that it is a specific human competence to develop normative standards on the basis of systematic actions or observations (Herrmann, Call, Hernandez-Lloreda, Hare, & Tomasello, 2007; Rakoczy, Warneken, & Tomasello, in press).

However, once both dimensions—the individual's cognitive dimension and the socio-normative dimension—are established, they constantly interact and re-model each other by bottom-up and top-down processes (see Fig. 2). For example, based on our low-level non-conceptual feelings about a certain action event we might modify our normative judgments, e.g. when being personally involved we might come to a normative judgment that is mainly carried by emotional involvement, but not by rational argument (Greene & Haidt, 2002; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001); or when lacking the capacity of emotionally assessing an action effect (e.g. due to a brain disease), we are impaired in forming out moral judgments (Damasio, 2003). Vice versa, based on our metarepresentational normative judgments and the standards of our society, we might change the judgments about our actions. If, for example, a certain action is socially highly accepted, we are more inclined to attribute agency to ourselves and maybe even develop a FoA. For example, when playing soccer, I hit the ball and thereby make a goal, I have the feeling that I just had the right reaction, whereas when my hitting the ball prevents it from going into the goal, I am more likely to have the feeling that the circumstances hindered my action.

5.3.2. Neurocognitive mechanism

To be capable to form an AoR, one has to combine the ability to build up meta-representations with the ability to form a special mental model of oneself and of the other person: One not only has to be able to register one's own and other people's mental states, but also to detect specifically those mental states that are part of the intentional action planning process, that is, the internal action plan, the internal representation of the short- and long-term consequences and the internal representation of the self (the self-model). Furthermore, one has to be able to understand action as associated with normative expectations and evaluations concerning others and concerning oneself. In short, a so-called Theory of Mind (ToM; Premack & Woodruff, 1978) is required for AoR.

There is still a lot of controversy about the best functional description of ToM: There are mainly three basic theories, namely the so-called theory-theories (Carruthers & Smith, 1996), the simulation theories (Davies & Stone, 1995; Dokic & Proust, 2002), and interaction theory (Gallagher, 2005). Whatever theory is on the right track, it is clear that a ToM involves the capacity to detect specific intentions (through the actions of others). This intention detection probably relies on the motor system's use of internal inverse and forward models. More specifically, the proposal is that other's intentions might be detected by activating internal sensorimotor representations (“internal models”) at a sub-threshold level when observing their actions (for more details see Bosbach, Cole, Prinz, & Knoblich, 2005; Hamilton, Wolpert, & Frith, 2004; Wolpert, Doya, & Kawato, 2003).

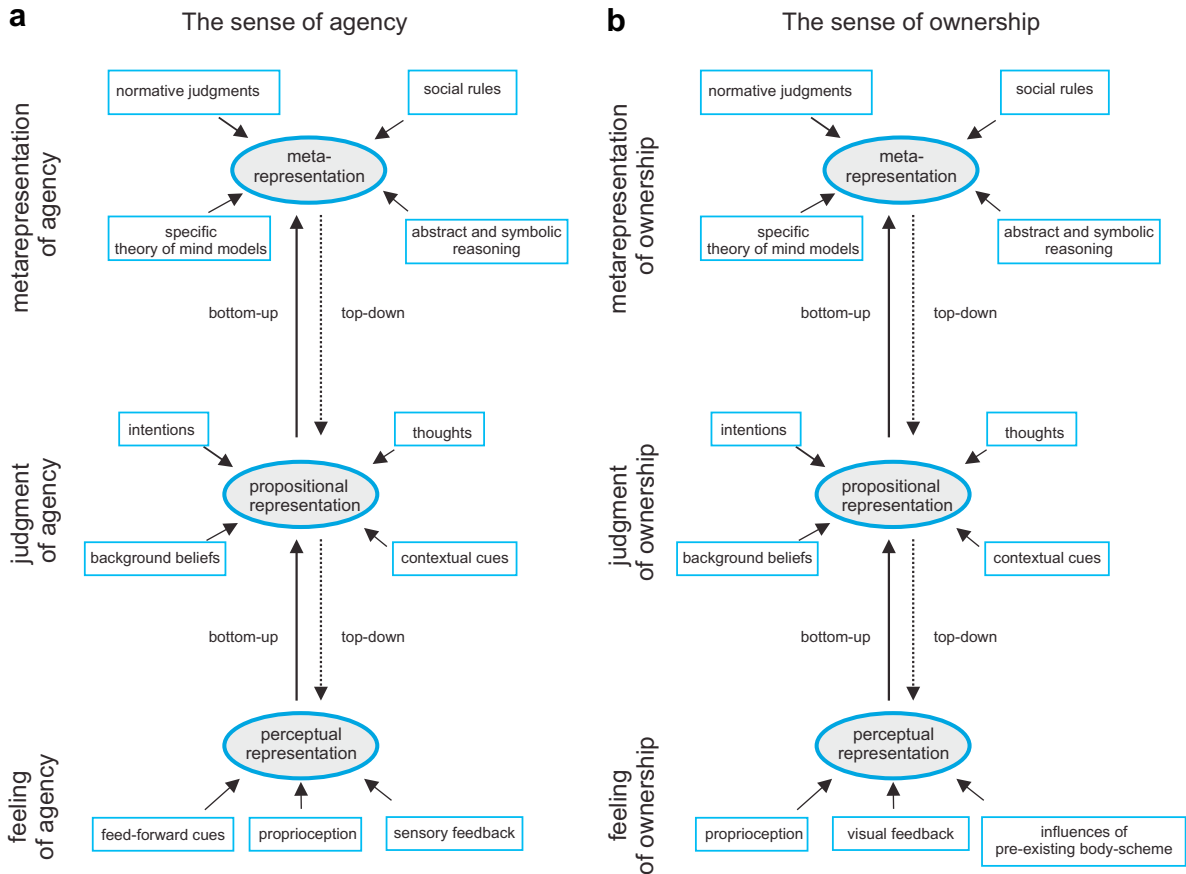


Fig. 2. Distinguishing different forms of bodily self-representations. The distinction between feeling, judgment, and meta-representation can be applied to both the sense of agency and the sense of ownership. A preliminary model suggests which different sensorimotor and cognitive authorship indicators enter each level of agency (a) and of ownership (b), respectively. For experimental investigation, either the relation and interaction between these six levels can be tested (e.g. in a 2 × 3-design) or each level can be tested per se (e.g. by modifying, adding or omitting one of the respective authorship indicators).

The competence to decode another's action intentions is not the only necessary element for the ascription of responsibility: The socio-normative dimension of ascription of responsibility demands an additional, more specific component, namely the understanding of norms and expectations and the capacity to act accordingly. Thus, further specific computational substrates are required:

- (1) A representation of a shared norms about what is expected,
- (2) The capacity to detect ongoing deviations from that norm and
- (3) To correct one's behavior accordingly, and
- (4) The capacity to do this not only from a first-person-perspective, but also from a third-person perspective, that is, from the point-of-view of a "neutral subject" that is maximally free of particular, individualistic motifs and interests.

This list (which roughly follows the distinction by [Montague & Lohrenz, 2007](#)) is obviously not complete, but it presents the fundamental requirements for a system in order to be capable of establishing an AoR. Moreover, these aspects have recently received first empirical corroboration with respect to their neural substrates (for review see [Montague & Lohrenz, 2007](#)) and, most interestingly with respect to clinical practice, offer a first preliminary tool to categorize different aspects in which socio-normative behavior might break-down in case of psychopathological disturbances or neurological lesions: Whereas a patient with some specific lesion in the prefrontal cortex might be able to represent a shared norm (criterion 1) and to detect ongoing deviations (criterion 2), but fails to correct his behavior accordingly (criterion 3), a patient with a lesion located in a slightly different prefrontal area might not even be able to represent the shared norm (criterion 1).

6. Dimensions of ownership

Like the SoA, also the SoO appears phenomenally as a central and uniform processing module, while in fact it is also a complex crossmodal phenomenon of largely heterogeneous functional and representational levels. These levels can be systematically identified by strictly analogous individuation criteria as were used for identifying levels of agency:

- (1) *Feeling of ownership* (FoO). On this primarily perceptual level, the non-conceptual, low-level representation of being the owner of one's own body parts leads to the FoO. Analogous to the FoA, the FoO results from a multimodal weighting of different sensory "ownership indicators", but—like the efferent components in the FoA—it also includes a non-sensory "top-down" component, namely a pre-existing internal body scheme which mediates and constrains sensory "bottom-up" inputs (Tsakiris & Haggard, 2005b) (see Fig. 2). In case of congruency between these ownership indicators (e.g. a match between proprioception and visual feedback), the experience of being the owner of one's own body parts is withheld from conscious processing and we experience bodily ownership by a rather diffuse feeling of a coherent, harmonious ongoing flow of bodily experiences. In case of incongruency between these indicators (e.g. a mismatch between proprioception, body schema and visual feedback), we experience our body parts as strange, peculiar or alien.⁹ On the level of feeling, a distinction between agency and ownership is present not only phenomenologically (an active movement feels different than a passive movement or non-moving state of a limb), but also with respect to the neurocognitive mechanism: In case of ownership, no internal prediction is issued and thus no comparator mechanism triggered (a fact which was overlooked by Gallagher, 2000).
- (2) *Judgment of ownership* (JoO). On this conceptual level, an explicit, interpretative judgment of being the owner of one's own body parts occurs. Analogous to the JoA, the JoO results from a weighting of different *cognitive* ownership indicators (for a comparison between FoA/JoA and FoO/JoO, see Fig. 2). Thus, on this level, the non-conceptual feeling of bodily ownership is further processed by conceptual capacities and belief stances to form an attribution of ownership. For example, a mismatch between different sensory ownership indicators triggers (i) a primary basic feeling of not being the owner of some body part and (ii) a second interpretative mechanism which looks for the best explanation, resulting in a specific belief formation about the belonging of the body part: I can either believe that I am the owner of the body part despite the mismatch or I can assume that someone or something other than me owns this body part. If an external ownership attribution occurs, it is moreover an open question to whom the action is attributed. Several illustrative examples for these processes will be given in the following section. There it will be revealed that the lack of the distinction between FoO and JoO has not only caused several confluences in the investigation of the neural substrates of the SoO (Tsakiris, Hesse, Boy, Haggard, & Fink, 2007), but also in the investigation of the neurocognitive processes and the neuropsychological disturbances of the SoO.
- (3) *Meta-representation of ownership* (MoO). At the meta-representational level, the notion of ownership is extended to objects which are not proper parts of my body but are mine in the sense of belongings.¹⁰ This extended notion of ownership requires an explicit theory and thereby depends on background beliefs and socio-cultural norms. Crucially, what belongs to me and what does not is—in the extended sense—a social question. Therefore, an adequate representation of the ownership of objects requires (1) that I represent the object as belonging to me, (2) that others represent the object as belonging to me, and (3) our mutual representation of these representations of the other, respectively. Since the latter is clearly a form of meta-representation, we can speak of the meta-representation of ownership. In this way, many things are represented as belongings of me by means of socio-normative attribution, thus depending heavily on beliefs about socio-normative relations. For example, according to the norms in Western Europe, kids belong to their parents while there exist also radical ideas (one may find, e.g. in Plato or in radical communism) according to which children are belonging to the whole society and not especially to their parents. There are great cultural differences concerning a social theory about what is part of me: For example, depending on the respective normative evaluation or social standards, I can consider not only my children, but also my whole family or even my social peer group or regional population to be "part of me". Thus, taken together, one's theory of body ownership is only one particular case of meta-representation of ownership—usually this meta-representation extends beyond the bodily self. As such a multidimensional bodily, psychological and social meta-representation, it forms the fundamental basis for our self-understanding.

Like the AoR, the neurocognitive requirements for building up a MoO presuppose the capacity to build up meta-representations with the ability to form a special mental model of oneself and of others, respectively. Additionally, one has to build up the computational substrates that are required to represent one's own and the other's ownership-attribution by social standards and normative evaluation. This means that one also has to be able to represent one's own and the other's norms about one's body and one's social environment plus the according socio-normative attribution mechanisms.

7. Empirical application of the frameworks

The multi-level frameworks of the SoA and the SoO can be used to inform various empirical disciplines in their investigation of agency and ownership, e.g. developmental psychology, psychiatry or cognitive neuroscience. Here, we take neuropsychology

⁹ Gallagher (2005) distinguishes between the body schema and the body image. We offer a more detailed distinction for both the body schema (which corresponds roughly to the registration of action-effect couplings and the FoO) and the body image (which contains both parts of what we call judgment of ownership and meta-representation of ownership).

¹⁰ On the one hand, also objects that are not proper parts of my body can be integrated into the body schema (e.g. the tennis racket) when using it extensively. This integration takes place on the level of FoO, but the object is not represented as a belonging (but wrongly as a body part). On the other hand, the extended notion of ownership is not confined to objects apart from my body: The question, whether a pregnant woman's body is her own in the sense that she is the only one to decide about abortion, for example, involves the extended notion of ownership for own body parts (we thank an anonymous reviewer for this point).

as an example to demonstrate the applicability and analytical benefit of these frameworks. In particular, it will be shown that they provide genuinely new and very differentiated explorative tools for analyzing and understanding disruptions of agency in psychopathological and neuropsychological cases. In turn, these cases will also provide empirical support for the multi-level frameworks by showing that their conceptual distinctions correspond to neuropsychological dissociations.

7.1. Disturbances of the sense of ownership

Current analyses of alien limb phenomena (Spence, 2002) are strongly confounded by the missing distinction between agency and ownership components. Elaborating on the original definition by Brion and Jedynak (1972), we here suggest that “alien limb phenomena” include only disturbances of the SoO, but not disturbances of the SoA. This characterization is based on the fact that—in contrast to the “anarchic hand”—alien limb phenomena do not per se entail any involuntary movements and thus no disruption of agency, but only a disturbance in the subjective experience of the ownership of one’s own limbs (for a roughly similar analysis see Gallagher & Veiver, 2004).

Following the thesis that the SoO is a complex crossmodal representation of different functional and representational levels, alien limb phenomena can be present on different levels of the SoO, namely on the level of FoO, JoO or MoO. Accordingly, the “alien hand syndrome” is an etiologically heterogeneous pathological disruption which might result from a deficit in the FoO in some patients, but from a deficit in the JoO in others. A genuine disturbance of the FoO consists in the perceptual experience of “alienness” (Bisiach & Geminiani, 1991, p. 20) of one’s own limbs—a perceptual experience which is, however, not further influenced by conceptual elaboration or involvement of belief states. Such a disturbance is described for patient LA-O who suffered from a unilateral neglect:

“On request, she admitted without hesitation that her left shoulder was part of her body and *inferentially* came to the same conclusion as regards her left arm and elbow, given, as she remarked, the evident continuity of those members. She was elusive about the forearm but insisted on denying ownership of the left hand, even when it had been passively placed on the right side of her trunk. She could not explain why her rings happened to be worn on the fingers of the alien hand.” (Bisiach & Geminiani, 1991, p. 32–33).

Although patient LA-O reports her disturbed SoO in a verbal form, it is very plausible that her experience results from a non-conceptual feeling of alienness and non-ownership. In contrast to this disturbance of the FoO, her JoO is still functioning adequately: (i) She is able to *inferentially* self-attribute at least some body parts (her left arm and elbow), and, even more importantly, (ii) she accepts her alien experience as alien, rather than forming a delusional ownership hypothesis about this experience. Both facts reveal a largely intact belief formation system with respect to the way of action rationalization and self-theorizing.

In contrast to this patient, other patients do not accept their peculiar body experience as a strange experience, but form a delusional ownership hypothesis about this experience and maintain it despite different stored encyclopedic knowledge about their body and despite the testimony of others. Consider, for example, the experience and attribution behavior of the following patient, who was suffering from a right-sided fronto-temporo-parietal infarction:

“Examiner: Whose arm is this?
A.R. (patient): It’s not mine.
Examiner: Whose is it?
A.R.: It’s my mother’s.
Examiner: How on earth does it happen to be here?
A.R.: I don’t know. I found it in my bed.
Examiner: How long has it been there?
A.R.: Since the first day.” (Bisiach, Rusconi, & Vallar, 1991)

This patient does not just have a perceptual experience of alienness, as patient LA-O, but actually transits from this abnormal experience to a formation of the (somatoparaphrenic) delusional belief that her arm belongs to someone else, namely her mother, thus indicating an additional deficit in the JoO. In fact, within the construction of the JoO, it might especially depend on the respective personal belief formation process whether an individual alien limb patient does actually develop (i) no delusional belief at all, (ii) the delusion of asomatognosia (the delusional belief that one’s own body part belongs *not to oneself*) or, by further positive determination of the delusional content, (iii) the delusion of somatoparaphrenia (the delusional belief that one’s body part belongs *to someone else*).

7.2. Disturbances of the sense of agency

Following the thesis that the SoA is a complex crossmodal representation of different functional and representational levels, also disturbances of the SoA can be present on different levels, namely on the level of FoA, JoA and AoR. This will be illustrated by two different disruptions of agency, namely “anarchic limb” and “anosognosia for one’s own hemiparesis”. While the former will illustrate primarily a failure of the JoA and partly even of the AoR, the latter will exemplify a disturbance primarily of the FoA and the JoA.

7.2.1. Anarchic limb

Anarchic limb phenomena encompass complex movements of the upper extremity which are not intended by the subject himself, although they are performed in a goal-directed and “intentional” way (i.e. they seem to serve a purpose) and with fairly adequate movement parameters (Della Sala, Marchetti, & Spinnler, 1991). In contrast to *alien limb*, anarchic limb clearly reflects a disturbance of the agency system, but not of the ownership system, in that the subject does not experience the limb as alien, but as part of his body, which is, however, controlled by intentions that are not part of the subject’s actual intentions. This claim is supported by the following patient report of experiencing an anarchic hand:

“[The patient’s] left hand would tenaciously grope for and grasp any nearby object, pick and pull at her clothes, and even grasp her throat during sleep [...]. She slept with the arm tied to prevent nocturnal misbehaviour. She never denied that her left arm and hand belonged to her, although she did refer to her limb as though it were an autonomous entity” (Banks et al., 1989, p. 456).

For this reason, one should strictly restrain from conflating disorders of agency and disorders of ownership by subsuming both of them under the term “alien hand” (Banks et al., 1989; Marchetti & Della Sala, 1998; Spence, 2002).

One might hold that “anarchic hand” is actually not a disturbance of the SoA, since patients correctly recognize that there is a discrepancy between the actual movement of their anarchic limb and their motor intentions (Frith et al., 2000). However, given our framework of the SoA, we can disentangle the different levels: Whereas there is no intention for the anarchic movement¹¹ and thus a discrepancy between the seeming intention of the movement and the intentions of the patient, the FoA does not occur (or, more precisely, **a feeling of non-agency occurs**), which is, in a certain way correct (although the seemingly intentional movement of the limb is a pathologic malfunctioning).¹² As a consequence of this situation, a feeling of the kind “This is not my action” is produced and it may be accompanied by a feeling of irritation. Given such a feeling, a rational way to come to a judgment concerning agency is the consideration that the activity of the limb is caused by its own, that it is (or is part of) a separate agent with intentions of its own. Hence, the agency-related problem lies on the level of the JoA, where a delusional belief is formed. Such inadequate beliefs might even have inadequate behavioral consequences. For example, one might *ascribe* (and not only perceive) intentions to one’s hand and come to the belief that the hand might thus be open to influence of verbal instructions:

“She often spoke to her own [affected] left hand, asking it to perform some movement, but the hand ‘did only what it wanted to do.’” (Trojano, Crisci, Lanzillo, Elefante, & Caruso, 1993, p. 215; cited after Spence, 2002).

Like the fact that judgments and beliefs about ownership cannot be (sufficiently) explained by feelings about ownership, beliefs about causation and intentional stances, i.e. about agency, can also not be explained by feelings about (non-)agency: Whereas the detection of a discrepancy between one’s actual movements and one’s motor intentions might lead to the perceptual experience of not being the initiator of the actual movement (FoA), it cannot explain why a person does or does not ascribe intentions to it (JoA). Again, this depends on an interpretative judgment that is formed by ad hoc rationalization that draws on contextual cues and belief states, in particular one’s personal background beliefs.

7.2.2. Anosognosia for one’s own hemiparesis

Patients exhibiting anosognosia for their own hemiparesis (AHP) most probably have a deficit in integrating the various sensorimotor authorship indicators, thus leading to an incorrect FoA (though it is still open to debate whether the deficit is primarily due to a deficit in integrating sensory feedback [“ignored feedback hypothesis”, (Levine, 1990)] or rather in integrating feed-forward motor intentions [“feed-forward theory”, Heilman, Barrett, & Adair, 1998]). However, the fact (i) that AHP patients do not simply accept their experience of having made a movement as false and (ii) that they cannot correct their belief about this experience despite of manifold evidence, clearly shows that they are not just impaired in their FoA but also in their JoA. Consequently, accounts which provide only an explanation for the disturbed FoA in AHP patients—like certain versions of the comparator model account (Frith et al., 2000)—are insufficient in that they cannot explain the disturbance of the JoA.

They fail even more in explaining why most of the AHP patients additionally develop abnormal beliefs towards the paretic limb, as, for example, *asomatognosia* (the delusional belief that a body part is not owned by themselves) or *somatoparaphrenia* (the delusional belief that a body part is owned by a certain other person).¹³ In fact, a recent analysis showed that AHP is not only behaviorally closely connected to abnormal beliefs about ownership and responsibility of the limb (92% of patients with AHP exhibit such abnormal beliefs), but also anatomically (the right posterior insula is commonly damaged in AHP patients with and without abnormal beliefs) (Baier & Karnath, 2008). Thus, due to the close connectedness of AHP and abnormal beliefs,

¹¹ The fact that the patient “perceives” an intention of the hand is not at all pathological: Every normal person (including the psychiatrist) “perceives” the intention of the hand in the very same way (as can be seen in the description of these cases, as e.g., “the anarchic hand prevents him from smoking”).

¹² Another interpretation would be that there is indeed an intention for the anarchic movement which is, for pathological reasons, detached from all other processing, such that neither (a) a feeling of ownership for the very *intention* arises (a topic which goes beyond the scope of this paper; see Vosgerau & Newen, 2007 for more details on this topic), nor (b) a feeling of agency can be established—instead, as already mentioned, a feeling of non-agency arises (which would be, under this interpretation, false).

¹³ Likewise, these accounts cannot explain why, e.g. some schizophrenia patients with delusions of control ascribe responsibility of certain actions to a specific other agent rather than to themselves or to nobody at all, cf. Synofzik et al., 2007.

theories about AHP cannot confine themselves to just explaining a possibly disturbed FoA, but, moreover, also have to explain the disturbed rationalization and belief modules.

8. Conclusion

Two central features of self-consciousness, the sense of ownership and the sense of agency, are based on complex mechanisms on several levels: A very basic level of sensorimotor integration, a non-conceptual level, a propositional level, and a meta-representational level. The multi-level framework presented here does not only give a detailed naturalistic account of these central aspects of self-consciousness but is also able to provide a comprehensive analysis and explanation of the different developmental stages of the SoO and the SoA and their specific pathological disturbances. Future research will refine the framework by more detailed conceptual and empirical investigation of the involved factors and processes at the different levels on the one hand, and of the interplay of the different levels on the other hand.

Acknowledgements

We thank two anonymous reviewers for their valuable comments and the “European Platform for Life Sciences, Mind Sciences, and the Humanities” sponsored by the Volkswagen Stiftung for their financial support.

References

- Baier, B., & Karnath, H. O. (2008). Tight link between our sense of limb ownership and self-awareness of actions. *Stroke*, 39(2), 486–488.
- Balslev, D., Cole, J., & Miall, R. C. (2007). Proprioception contributes to the sense of agency during visual observation of hand movements: Evidence from temporal judgments of action. *Journal of Cognitive Neuroscience*, 19(9), 1535–1541.
- Banks, G., Short, P., Martinez, J., Latchaw, R., Ratcliff, G., & Boller, F. (1989). The alien hand syndrome. Clinical and postmortem findings. *Archives of Neurology*, 46, 456–459.
- Bell, C. C. (2001). Memory-based expectations in electrosensory systems. *Current Opinion in Neurobiology*, 11, 481–487.
- Bermúdez, J. L. (1998). *The paradox of self-consciousness*. Cambridge: MIT Press.
- Bisiach, E., & Geminiani, G. (1991). Anosognosia related to hemiplegia and hemianopia. In G. P. Prigatano & D. L. Schacter (Eds.), *Awareness of deficit after brain injury* (pp. 17–39). New York: Oxford University Press.
- Bisiach, E., Rusconi, M. L., & Vallar, G. (1991). Remission of somatoparaphrenic delusion through vestibular stimulation. *Neuropsychologia*, 29(10), 1029–1031.
- Bosbach, S., Cole, J., Prinz, W., & Knoblich, G. (2005). Inferring another's expectation from action: The role of peripheral sensation. *Nature Neuroscience*, 8(10), 1295–1297.
- Brion, S., & Jedynak, C. P. (1972). Troubles du transfert interhemispherique. A propos de trois observations de tumeurs du corps calleux. Le signe de la main étrangere. *Revisa de Neurologia (Paris)*, 126, 257–266.
- Carruthers, P., & Smith, P. K. (Eds.). (1996). *Theories of theories of mind*. Cambridge: Cambridge University Press.
- Damasio, A. (2000). *The feeling of what happens: Body and emotion in the making of consciousness*. Fort Washington: Harvest.
- Damasio, A. R. (2003). *Looking for spinoza: Joy, sorrow, and the feeling brain*. New York: Blackwell.
- Davies, A. R., & Stone, T. (Eds.). (1995). *Mental simulation*. Oxford: Blackwell.
- Davies, M., Coltheart, M., Langdon, R., & Breen, N. (2001). Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, Psychology*, 8(2–3), 133–158.
- de Vignemont, F., & Fournieret, P. (2004). The sense of agency: A philosophical and empirical review of the “Who” system. *Consciousness and Cognition*, 13(1), 1–19.
- Della Sala, S., Marchetti, C., & Spinnler, H. (1991). Right-sided anarchic (alien) hand: A longitudinal study. *Neuropsychologia*, 29(11), 1113–1127.
- Dokic, J., & Proust, J. (Eds.). (2002). *Simulation and knowledge of action*. Amsterdam: John Benjamins.
- Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 229–240.
- Farrer, C., Franck, N., Paillard, J., & Jeannerod, M. (2003). The role of proprioception in action recognition. *Consciousness and Cognition*, 12(4), 609–619.
- Feinberg, I. (1978). Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophr Bull*, 4, 636–640.
- Fournieret, P., Paillard, J., Lamarre, Y., Cole, J., & Jeannerod, M. (2002). Lack of conscious recognition of one's own actions in a haptically deafferented patient. *Neuroreport*, 13(4), 541–547.
- Frith, C. (1992). *The cognitive neuropsychology of schizophrenia*. Hilsdale: Erlbaum.
- Frith, C. (2005). The self in action: Lessons from delusions of control. *Consciousness and Cognition*, 14(4), 752–770.
- Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 355(1404), 1771–1788.
- Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4, 14–21.
- Gallagher, S., & Væver, M. (2004). Disorders of embodiment. In J. Radden (Ed.), *The philosophy of psychiatry: A companion* (pp. 118–132). Oxford: Oxford University Press.
- Gallagher, S. (2005). *How the body shapes the mind?* Oxford: OUP.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108.
- Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6, 517–523.
- Haarmeier, T., Bunjes, F., Lindner, A., Berret, E., & Thier, P. (2001). Optimizing visual motion perception during eye movements. *Neuron*, 32(3), 527–535.
- Hamilton, A., Wolpert, D., & Frith, U. (2004). Your own action influences how you perceive another person's action. *Current Biology*, 14(6), 493–498.
- Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.
- Heilman, K. M., Barrett, A. M., & Adair, J. C. (1998). Possible mechanisms of anosognosia: A defect in self-awareness. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 353(1377), 1903–1909.
- Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, 317(5843), 1360–1366.
- Hommel, B., & Elsner, B. (in press). Acquisition, representation, and control of action. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *The Psychology of Action*. Oxford: Oxford University Press.
- Ito, M. (2005). Bases and implications of learning in the cerebellum—adaptive control and internal model mechanism. *Progress in Brain Research*, 148, 95–109.
- Legrand, D. (2007). Subjectivity and the body: Introducing basic forms of self-consciousness. *Consciousness and Cognition*, 16(3), 577–582.

- Levine, D. N. (1990). Unawareness of visual and sensorimotor defects: A hypothesis. *Brain and Cognition*, 13(2), 233–281.
- Lindner, A., Haarmeier, T., Erb, M., Grodd, W., & Thier, P. (2006). Cerebrocerebellar circuits for the perceptual cancellation of eye-movement-induced retinal image motion. *Journal of Cognitive Neuroscience*, 18(11), 1899–1912.
- Marchetti, C., & Della Sala, S. (1998). Disentangling the alien and anarchic hand. *Cognitive Neuropsychiatry*, 3, 191–208.
- Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. Cambridge: MIT.
- Montague, P. R., & Lohrenz, T. (2007). To detect and correct: Norm violations and their enforcement. *Neuron*, 56(1), 14–18.
- Newen, A., & Vogeley, K. (2003). Self-representation: Searching for a neural signature of self-consciousness. *Consciousness and Cognition*, 12(4), 529–543.
- Newen, A., & Bartels, A. (2007). Animal minds and the possession of concepts. *Philosophical Psychology*, 20(3), 283–308.
- Newen, A., & Vogeley, K. (2007). Menschliches Selbstbewusstsein und die Fähigkeit zur Zuschreibung von Einstellungen (Theory of Mind). In H. Förstl (Ed.), *Theory of Mind. Neurobiologie und Psychologie sozialen Verhaltens* (pp. 100–116). Heidelberg: Springer.
- Poulet, J. F., & Hedwig, B. (2002). A corollary discharge maintains auditory sensitivity during sound production. *Nature*, 418(6900), 872–876.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a “theory of mind”? *Behavioural Brain Research*, 4, 515–526.
- Rakoczy, H., Warneken, F., & Tomasello, M. (in press). The sources of normativity: Young children’s awareness of the normative structure of games. *Developmental Psychology*.
- Rochat, P., & Striano, T. (1999). Emerging self-exploration by 2-month-old infants. *Developmental Science*, 2, 206–218.
- Spence, S. A. (2002). Alien motor phenomena: A window on to agency. *Cognitive Neuropsychiatry*, 7(3), 211–220.
- Synofzik, M., Thier, P., & Lindner, A. (2006). Internalizing Agency of Self-Action: Perception of One’s Own Hand Movements Depends on an Adaptable Prediction About the Sensory Action Outcome. *Journal of Neurophysiology*, 96(3), 1592–1601.
- Synofzik, M., Vosgerau, G., & Newen, A. (2007). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Trojano, L., Crisci, C., Lanzillo, B., Elefante, R., & Caruso, G. (1993). How many alien hand syndromes? Follow-up of a case. *Neurology*, 43(12), 2710–2712.
- Tsakiris, M., & Haggard, P. (2005a). Experimenting with the acting self. *Cognitive Neuropsychology*, 22(3/4), 387–407.
- Tsakiris, M., & Haggard, P. (2005b). The rubber hand illusion revisited: Visuotactile integration and self-attribution. *Journal of Experimental Psychology. Human Perception and Performance*, 31(1), 80–91.
- Tsakiris, M., Hesse, M. D., Boy, C., Haggard, P., & Fink, G. R. (2007). Neural signatures of body ownership: A sensory network for bodily self-consciousness. *Cerebral Cortex*, 17(10), 2235–2244.
- Vosgerau, G. (2007). Conceptuality in spatial representations. *Philosophical Psychology*, 20(3), 349–365.
- Vosgerau, G., & Newen, A. (2007). Thoughts, Motor actions, and the Self. *Mind and Language*, 22, 22–43.
- Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, 11(18), R729–R732.
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358(1431), 593–602.
- Zinck, A., & Newen, A. (2007). Classifying emotion: A developmental account. *Syntheses*, 49.