

---

# Empirische Perspektiven aus Sicht der Selbstmodell-Theorie der Subjektivität: Eine Kurzdarstellung mit Beispielen\*

THOMAS METZINGER  
*Johannes Gutenberg-Universität Mainz*

**Streszczenie.** *Dieser Beitrag bietet eine Kurzdarstellung der Selbstmodell-Theorie der Subjektivität (SMT; Metzinger, 2003a) für empirische Forscher. Anhand von empirischen Beispielen aus verschiedenen Wissenschaftsdisziplinen werden einige Mechanismen erörtert, durch die in einem physisch realisierten Informationsverarbeitungssystem wie dem menschlichen Gehirn Selbstbewußtsein entstehen könnte. Der Beitrag stellt zwei zentrale Begriffe vor: das »phänomenale Selbstmodell« (PSM) und das »phänomenale Modell der Intentionalitätsrelation« (PMIR). Dabei wird eine repräsentationalistische Analyse des bewußten Selbst und der Entstehung einer Erste-Person-Perspektive skizziert.*

**Schlüsselwörter:** *Bewußtsein; Selbstbewußtsein; Erste-Person-Perspektive; Meinigkeit; Agentivität; Selbstmodell; phänomenale Transparenz; Phantomglieder; Robotik; Gummihand-Illusion; Ganzkörper-Illusion; außerkörperliche Erfahrung; Evolution des Werkzeuggebrauchs*

---

\*Dieser Text ist eine stark erweiterte, aktualisierte und überarbeitete Version eines Artikels, der 2008 in der Zeitschrift *Progress in Brain Research* erschienen ist. Ich danke Sarah Raphael für ihre Hilfe bei der Übersetzung einer ersten Rohfassung dieses Textes aus dem Englischen und Jennifer M. Windt für eine Vielzahl von kritischen Kommentaren und Verbesserungsvorschlägen. Ich bin auch Ingrid Götsch zu Dank verpflichtet, die mich in der Schlußphase auf zuverlässige und professionelle Weise unterstützt hat. Für ihre sorgfältige Hilfe bei der Schlußkorrektur danke ich Hannes Boelsen, Regina Fabry und Lisa Quadt.

### Der vierte Schritt: Das körperliche Selbst als eine funktionale Verankerung des phänomenalen Raums

Ich habe oben die Unterscheidung zwischen einer repräsentationalen und einer funktionalen Analyse der Erste-Person-Perspektive gemacht. Das zentrale theoretische Problem auf der funktionalen Beschreibungsebene kann mit folgender Frage zusammengefaßt werden: Was genau ist der spezifische Unterschied zwischen dem PSM und anderen phänomenalen Modellen, die gegenwärtig im System aktiv sind? Gibt es eine charakteristische kausale Signatur des PSM? Durch welche *funktionale Eigenschaft* wird es zum stabilen Zentrum des phänomenalen repräsentationalen Raums?

Hier ist meine vorläufige Antwort: Das Selbstmodell ist die einzige repräsentationale Struktur, die durch eine *kontinuierliche Quelle intern generierten Inputs* im Gehirn verankert ist. Nennen wir dies die »Hypothese der permanenten kausalen Verbindung«. Immer dann, wenn bewußtes Erleben entsteht (also wenn ein stabiles, integriertes Modell der Wirklichkeit aktiviert wird), existiert auch diese kontinuierliche Quelle interozeptiven Inputs<sup>1</sup>. Das menschliche Selbstmodell besitzt eine dauerhafte und maximal direkte kausale Verknüpfung im Gehirn, und zwar durch die Eigenwahrnehmung des Körpers. Es hat Teile, die durch *permanente* Formen der Verarbeitung von *permanenten* Formen selbstgenerierten Inputs, zum Beispiel auch aus niedrigstufiger, homöostatischer Autoregulation, realisiert werden. Um diesen grundlegenden Gedanken anders auszudrücken: Der Körper ist das einzige Wahrnehmungsobjekt, vor dem das Gehirn nie weglaufen kann. Es gibt eine Reihe offensichtlicher Kandidaten für Quellen hoher Invarianz, zum Beispiel die folgenden vier Typen von intern erzeugter Information, die während Episoden des bewußten Erlebens ein dauerhaftes funktionales Bindeglied zwischen dem PSM und seiner körperlichen Basis im Gehirn erzeugen:

- Input aus dem Vestibulärorgan: der Gleichgewichtssinn.
- Input aus dem autonom aktiven, invarianten Teil des Körperschemas: Das andauernde »Hintergrundgefühl« im räumlichen Körpermodell, das unabhängig ist von externem Input, z.B. durch Bewegungswahrnehmung.

---

<sup>1</sup>Die Bedeutung von Interozeption für das Selbstbewußtsein wird in einer ganzen Reihe von neueren Studien von Manos Tsakiris und Kollegen untersucht; siehe etwa Tsakiris et al., 2011, Ainley et al., 2012. Ein interessantes komputationales Modell ist Seth, Suzuki & Critchley, 2012.

- Input aus Sensoren der Eingeweide, aber auch aus den Blutgefäßen, zum Beispiel aus den kardiovaskulären Mechanosensoren: »Bauchgefühle« und somatoviszzerale Formen der Selbstrepräsentation.
- Input aus bestimmten Teilen des oberen Hirnstamms und des Hypothalamus: Hintergrundemotionen und -stimmungen, die in der kontinuierlichen homöostatischen Selbstregulation des »internen Milieus« verankert sind, also der biochemischen Landschaft in unserem Blut.

Philosophisch gesehen sind es nicht so sehr die neurobiologischen Details, die entscheidend sind, sondern vielmehr die plausible Annahme, daß es einen bestimmten Teil des menschlichen Selbstmodells gibt, der sich durch einen hohen Grad an Stimuluskorrelation auszeichnet und ausschließlich auf *intern* generierter Information beruht. Diese Ebene des PSM ist direkt und permanent in Stimuli aus dem Inneren des Körpers verankert. Diese Tatsache ist phänomenologisch relevant, weil sie den entscheidenden Beitrag zur quasi-cartesischen Phänomenologie der Substantialität (»Ich bin eine ontologisch autonome Entität, die sich selbst ›im Sein halten‹ kann«) und der Selbstgewisheit des Subjekts (»Ich weiß, daß ich *weiß*, daß ich selbst existiere«) leistet. Sie ist aber auch erkenntnistheoretisch relevant, weil sie auf der funktionalen Ebene die garantierte Referenz erzeugt, von der oben die Rede war. Erinnern Sie sich noch an die Patientin AZ aus dem zweiten Beispiel? Der schwächere Grad an phänomenaler »Lebhaftigkeit« oder »Echtheit« in ihren Phantomgliedern könnte eben gerade die Abwesenheit von permanenter *bottom-up*-Stimulation widerspiegeln, die unter normalen Umständen von existierenden physischen Gliedern verursacht wird. Marcel Kinsbourne hat in diesem Zusammenhang von einem »background 'buzz' of somatosensory input« (Kinsbourne, 1995, S. 217) gesprochen. Um die Phänomenologie, die auf der repräsentationalen Beschreibungsebene an diesem »puren Körperlichkeitsgefühl« beteiligt ist, zu erfassen, möchte ich zwischen Selbstpräsentation und Selbstrepräsentation unterscheiden<sup>2</sup>. Phänomenologisch gesehen bezieht sich

---

<sup>2</sup>Zu einer ausführlichen theoretischen Behandlung dieses Themas und zahlreichen jüngeren empirischen Ergebnissen zum Körper als Anker des bewußten Erlebens vgl. Damasio (1999). Antonio Damasio verwendet den Begriff eines Kernselbst (*core self*) und an anderer Stelle (Metzinger, 1993, S. 156ff; Metzinger, 2003a, Abschnitt 5.4) habe ich den technischen Begriff der »phänomenalen Selbstpräsentation« (gegenüber der Selbstrepräsentation) eingeführt. Auf der Ebene der Körperrepräsentation ist die Selbstrepräsentation das, was AZ in ihren Phantomgliedern fehlt, während Selbstrepräsentation das ist, was sie tatsächlich besitzt — obwohl dies offen-

der erste Begriff auf das rein sensorische Gefühl der körperlichen Präsenz, welches interessanterweise einhergeht mit einer subjektiven Wahrnehmung zeitlicher Unmittelbarkeit und der auf Erfahrung beruhenden Sicherheit, eine direkte, nicht inferentielle Selbstkenntnis zu besitzen. Was genau ist diese tiefste Schicht des phänomenalen Selbst? Warum ist es der Ursprung der Erste-Person-Perspektive? Meine Hypothese besagt, daß nach der raumzeitlichen Selbstlokalisierung die konstante Selbstorganisationsaktivität derjenigen Regionen des körperlichen Selbst, die von externem Input unabhängig sind, das funktionale Zentrum des phänomenalen Repräsentationsraums bildet.

Als ein erstes Beispiel dafür, was man unter einem Selbstmodell verstehen kann, hatte ich Ramachandrans Experiment zur Mobilisierung eines gelähmten Phantomglieds vorgestellt. Eine Selbst*präsentation* ist genau der Teil des Phantomglieds, der unabhängig von Bewegungen immer bewußt bleibt. Wenn man *diesen* Teil verliert, verliert man auch das subjektive Erleben körperlicher Präsenz — man wird zu einem »körperlosen Wesen«<sup>3</sup> Jedoch könnte es sogar weitere, allgemeinere empirische Perspektiven geben, aus denen das Selbstmodell zwangsläufig mit der Basisaktivität des Gehirns, wie man sie im Ruhezustand beobachten kann, verbunden ist (vgl. Raichle et al., 2001; Gusnard, 2005).

### **Der fünfte Schritt: Autoepistemische Geschlossenheit — Die Transparenzannahme und das naiv-realistische Selbstmißverständnis**

Das zentrale theoretische Problem auf der repräsentationalen Beschreibungsebene besteht darin, daß man mir leicht vorwerfen könnte, ich würde mit der Einführung des Begriffs »Selbstmodell« das eigentliche Problem terminologisch umgehen. Die zentrale Eigenschaft des Selbstbewußtseins — die Tatsache, daß wir uns immer als Selbst erleben, aber nicht als Modell — scheint durch die SMT somit nicht

---

sichtlich auch eine Form der *Fehl*repräsentation ist, da der Referent dieser Repräsentation nie existiert hat.

<sup>3</sup>Auch hierzu existieren die entsprechenden phänomenologischen Zustandsklassen. In Metzinger (1993) und Metzinger (1997) habe ich Oliver Sacks' Beispiel der »körperlosen Lady« besprochen. Vgl. in diesem Zusammenhang auch den berühmten Fall von Ian Waterman, der in Metzinger (2003a) diskutiert wird. Interessant ist jedoch auch die neuere Traumforschung, die zeigt, wie es möglich ist, daß das phänomenale Subjekt stabil sein kann, sich aber trotzdem lediglich als ausdehnungsloser Punkt in einem räumlichen Bezugsrahmen selbst lokalisiert, Windt, 2010, 2014.

erklärt zu werden. Aber zum einen ist ein Selbstmodell natürlich nicht ein statischer repräsentationaler Träger, das Modell eines mysteriösen Einzeldings, das wir dann *das Selbst* nennen. Es ist ein kontinuierlicher und auf das betreffende System selbst gerichteter Prozeß, der globale Eigenschaften des Organismus in ihrer zeitlichen Entwicklung verfolgt und überwacht. Zum anderen scheint es keine notwendige Verbindung zwischen den diesen Vorgang charakterisierenden funktionalen und repräsentationalen Eigenschaften einerseits und den *phänomenalen* Zieleigenschaften der »Meinigkeit«, »präreflexiven Selbstheit« und »Perspektivität« andererseits zu geben — zumindest gewissen modalen Intuitionen zufolge. Manche denken: All das könnte sich durchaus ereignen, ohne daß es zu Bewußtsein, zur Entstehung eines echten phänomenalen Selbst oder einer subjektiven Innenperspektive kommt. Solange der Begriff »Bewußtsein« noch so extrem inhaltsleer ist, wie dies in der gegenwärtigen Debatte der Fall ist, scheinen Zombies zumindest logisch möglich zu bleiben: Es ist denkbar, daß biologische Informationsverarbeitungssysteme einen Darstellungsraum entwickeln und erfolgreich anwenden können, in dessen Zentrum ein Selbstmodell steht, *ohne* dabei auch phänomenales Selbstbewußtsein zu entwickeln. Etwas interessanter ist der mögliche Einwand, daß die Entstehung der spezifischen phänomenalen Qualität der *Selbstheit* — sogar dann, wenn man *die phänomenologische Ebene als gegeben voraussetzt*, d.h. sogar in einem System, das schon bewußt ist — nicht offensichtlich oder selbstverständlich ist. Was genau würde, mit begrifflicher Notwendigkeit, ein echtes Ego erzeugen? Ein »Selbstmodell« ist noch lange kein Selbst, sondern nur eine Repräsentation des Systems als einer Ganzheit — es ist nicht mehr als ein integriertes *Systemmodell*. Wenn die funktionale Eigenschaft der Zentriertheit und die repräsentationale Eigenschaft der Selbstmodellierung zu der phänomenalen Zieleigenschaft eines echten, bewußt erlebten »Ichgefühls« führen sollen, dann muß das bewußte Systemmodell zu einem phänomenalen Selbst werden. Die entscheidende philosophische Frage ist deshalb die folgende: Wie führt die Existenz eines funktional zentrierten Darstellungsraums notwendigerweise zu der Entstehung eines bewußten Selbst und dann zu dem, was wir gemeinhin eine phänomenale Erste-Person-Perspektive nennen? Mit anderen Worten: Wie wird das Systemmodell zu einem *Selbstmodell*?

Ein genuines, bewußtes Selbst — so lautet meine Antwort — entsteht immer genau in dem Moment, in dem das System das Selbstmodell, das es gerade auf der Ebene des bewußten Erlebens generiert, nicht mehr *als* Modell erleben kann. Wie kommt man also von der funktionalen Eigenschaft der »Zentriertheit« und der repräsentationalen Eigenschaft der »Selbstmodellierung« zu der phänomenalen Eigenschaft der »prä-

flexiven Selbstvertrautheit«? Die Lösung hängt mit dem zusammen, was Philosophen »phänomenale Transparenz« nennen (eine kurze Erläuterung des Begriffs findet sich in Metzinger, 2003c; Metzinger, 2003b ist der deutsche Vorläufer). Die vom System generierten repräsentationalen Zustände sind *transparent*, d.h. sie stellen die Tatsache, daß sie Modelle sind, nicht mehr auf der Ebene ihres Inhalts dar. Folglich — und dies ist zunächst nur eine visuell-phänomenologische Metapher, die vor über hundert Jahren von dem britischen Philosophen George Edward Moore eingeführt wurde — schaut das System durch seine eigenen repräsentationalen Strukturen »hindurch«, als ob es sich in direktem und unmittelbarem Kontakt mit ihrem Inhalt befände. Dies ist wohlbeachtet lediglich eine Aussage über die *Phänomenologie* des Systems. Es ist keine erkenntnistheoretische Aussage über den Besitz von Wissen: Man kann vollkommen verblendet sein und nichts oder sehr wenig über die Realität wissen (oder auch über den eigenen Geist) und sich dennoch der Phänomenologie der Gewißheit oder *des Wissens, daß man weiß*, erfreuen. Phänomenale Transparenz ist also nicht *epistemische* Transparenz, oder Descartes' klassische — und mittlerweile empirisch widerlegte — Vorstellung, daß wir uns über den Inhalt unseres eigenen Geistes nicht täuschen können. Transparenz ist in diesem Zusammenhang ausschließlich als eine Eigenschaft des *bewußten* Zustands definiert. Unbewußte Zustände sind weder transparent noch opak. Phänomenale Transparenz steht ebenso wenig in direkter Verbindung mit dem dritten technischen Begriff in der Philosophie, nämlich mit »referentieller Transparenz«. Wesen, die weder eine Sprache noch das begriffliche Denken beherrschen, könnten deshalb durchaus phänomenal transparente Zustände haben, ebenso wie Systeme, die einer fundamentalen Selbsttäuschung unterliegen. Der naive Realismus ist hier keine Überzeugung, kein spezifischer semantischer Kontext oder eine kognitive, epistemische Einstellung, sondern das, was diese überhaupt erst ermöglicht: eine Eigenschaft des phänomenalen Erlebens selbst.

Ich vertrete zwei Hypothesen bezüglich der mikrofunktionalen Basis und der evolutionären Entstehungsgeschichte transparenter phänomenaler Zustände. Erstens: In einem sehr kleinen Zeitfenster werden die fraglichen neuronalen Datenstrukturen so schnell und zuverlässig aktiviert, daß das System sie nicht mehr als solche erkennen kann, z.B. wegen der vergleichsweise langsamen zeitlichen Auflösung *metarepräsentationaler* Funktionen oder weil sie zu einem stabilen, Bayes-optimalen Modell gehören, für das der Vorhersage-Irrtum bereits stark minimiert wurde (vgl. Howhy 2013, Friston 2010). Introspektiv ist der Konstruktionsprozeß unsichtbar. Zweitens: Durch ein viel größeres Zeitfenster betrachtet gab es anscheinend keinen evolutionären Selektionsdruck auf

die entsprechenden Teile unserer funktionalen Architektur. Der naive Realismus war für biologische Systeme wie uns eine funktional adäquate Hintergrundannahme. Wir mußten wissen: »Vorsicht! Es ist ein Wolf in der Nähe!«, aber nicht: »Eine Wolf-Repräsentation ist momentan in meinem Gehirn aktiv!«

Phänomenale Transparenz hat also doch etwas mit Wissen zu tun (und nicht nur mit dem subjektiven Erleben *per se*): Es ist eine besondere Form der Dunkelheit. Es ist ein Mangel an Wissen. Epistemologisch gesehen ist sie ein impliziter Mangel an Selbstkenntnis, an Wissen über die funktionale Tiefenstruktur des eigenen Geistes, der aber selbst wiederum nicht wieder explizit repräsentiert wird. Wie Franz Brentano ([1874] 1973, S. 165f) und heute Daniel Dennett (1991, S. 359) deutlich gemacht haben, ist die Repräsentation von Abwesenheit nicht dasselbe wie Abwesenheit von Repräsentation. In transparenten Zuständen existiert keine Repräsentation vorhergehender Verarbeitungsstufen oder dynamisch instabiler Zustände. In der Phänomenologie des visuellen Erlebens bedeutet Transparenz, daß wir etwas nicht sehen können. Phänomenale Transparenz *im Allgemeinen* dagegen bedeutet, daß der Repräsentationscharakter des Gehalts des bewußten Erlebens selbst für das subjektive Erleben nicht zugänglich ist: Wir können nicht mehr sehen, daß es sich hier um eine *Vorhersage* handelt, um ein sehr gutes inneres Modell, dessen Funktion darin besteht, Überraschungen in der Zukunft zu vermeiden. Diese Analyse bezieht sich auf alle Sinnesmodalitäten, insbesondere auf das integrierte phänomenale Modell der Welt als Ganzer. Weil das *Mittel* der Darstellung als solches nicht dargestellt werden kann, wird das erlebende System notwendigerweise in einen naiven Realismus verstrickt; es erlebt sich selbst in direktem Kontakt mit dem Inhalt seines eigenen bewußten Erlebens. Was es nicht erleben kann, ist die Tatsache, daß sein Erleben immer in einem *Medium* stattfindet — und eben dies ist gemeint, wenn wir von der »Unmittelbarkeit« des phänomenalen Bewußtseins sprechen. In einer vollständig transparenten Repräsentation können die Mechanismen, die zu ihrer Aktivierung führen, und die Tatsache, daß es einen konkreten inneren Zustand gibt, der seinen Gehalt trägt, introspektiv nicht mehr erkannt werden. Im Jargon der traditionellen analytischen Philosophie des Geistes gesprochen: »Nur Inhaltseigenschaften sind introspektiv verfügbar, Eigenschaften des repräsentationalen Trägers sind unzugänglich.« Die Phänomenologie der Transparenz ist daher die Phänomenologie des naiven Realismus.

Viele phänomenale Repräsentationen sind transparent, weil ihr Inhalt und vor allem dessen Existenz in allen möglichen Kontexten festzustehen scheint: Das Buch, das Sie jetzt in den Händen halten, wird

dem subjektiven Erleben nach immer nur diese Buch bleiben — unabhängig davon, wie sich die äußerliche Wahrnehmungssituation verändert. Sie erleben niemals, daß ein »aktiver Objektemulator« in Ihrem Gehirn gerade in Ihr globales Realitätsmodell integriert wird. Sie erleben lediglich den *Inhalt* des zugrunde liegenden Repräsentationsvorgangs: das *Buch* als Ihnen jetzt und hier anstrengungslos gegebenes Buch. Die beste Art und Weise, sich den Begriff der Transparenz klarzumachen, besteht darin, zunächst ganz traditionell zwischen dem Vehikel und dem Gehalt einer Repräsentation zu unterscheiden, zwischen repräsentationalem Träger und repräsentationalem Inhalt (vgl. dazu auch Dretske, 1998, S. 45ff).

Der repräsentationale Träger Ihres bewußten Erlebens ist ein bestimmter Prozeß im Gehirn. Dieser Prozeß — der in keiner Weise etwas »Buchhaftes« an sich hat — wird nicht bewußt erlebt; er ist transparent in dem Sinne, daß Sie, phänomenologisch gesehen, durch ihn hindurch schauen. Worauf Sie schauen ist sein repräsentationaler Inhalt, die durch die eigene Sinneswahrnehmung vermittelte Existenz eines Buches, hier und jetzt. Aus der begrifflichen Außenperspektive betrachtet ist der Inhalt eine abstrakte Eigenschaft des konkreten repräsentationalen Zustands im Gehirn. Wenn der repräsentationale Träger ein gut und zuverlässig funktionierendes Instrument zur Wissensgewinnung ist, erlaubt er Ihnen dank seiner Transparenz, »durch ihn hindurch« direkt auf die Welt, auf das Buch in Ihrer Hand zu schauen. Er macht die von ihm getragene Information global verfügbar, ohne daß Sie — die Person als Ganzes — sich darum kümmern müssen, *wie* das geschieht. Das Besondere an den meisten phänomenalen Repräsentationen ist, daß Sie diesen Inhalt aus der Innenperspektive auch dann als maximal *konkret* und unzweideutig, als direkt und unmittelbar gegeben erleben, wenn das fragliche Objekt — das Buch in Ihrer Hand — gar nicht existiert, sondern nur eine Halluzination ist. Phänomenale Repräsentationen scheinen genau die Art von Repräsentationen zu sein, bei denen wir repräsentationalen Gehalt und repräsentationalen Träger auf der Ebene des subjektiven Erlebens nicht unterscheiden können.

Es gibt natürlich Gegenbeispiele und diese sind hilfreich, um den Begriff der »Transparenz« noch besser zu verdeutlichen. *Opake* phänomenale Repräsentationen entstehen zum Beispiel dann, wenn die Information, daß es sich bei ihrem Inhalt um das Resultat eines inneren Darstellungsvorgangs handelt, plötzlich global verfügbar wird. Wenn Sie plötzlich entdecken, daß es das Buch in Ihrer Hand in Wirklichkeit nicht gibt, wird die Halluzination zu einer Pseudohalluzination. Die Information, daß Sie nicht auf die Welt, sondern »auf« einen aktiven repräsentationalen Zustand schauen, der im Moment allem An-



schein nach nicht als ein zuverlässiges Instrument zur Wissensgewinnung fungiert, ist jetzt auch auf der Ebene des subjektiven Erlebens selbst verfügbar. Der phänomenale Buchzustand wird undurchsichtig. Die *sensorische* Transparenz geht verloren. Ihnen wird die Tatsache bewußt, daß Ihre Wahrnehmungen durch Sinnesorgane erzeugt werden und daß diese Organe nicht in allen Situationen absolut zuverlässig funktionieren. Sie erleben das Buch jetzt nicht nur plötzlich als Repräsentation, sondern erleben es als *Fehl*repräsentation.

Nehmen wir weiter an, daß Sie jetzt sogar plötzlich entdecken, daß nicht nur die Buchwahrnehmung, sondern auch Ihr gesamtes philosophisches Nachdenken über das Problem des Bewußtseins in einem Traum stattfindet. Dann wird dieser Traum zu einem Klartraum (warum Klarträume eine philosophisch relevante Klasse von Bewußtseinszuständen sind, erläutere ich in Metzinger, 2003a, Abschnitt 7.2.5; mehr zu diesem Thema ist zu finden in Windt und Metzinger, 2007; Noreika et al., 2010; insbesondere in Windt, 2014). Die Tatsache, daß Sie momentan nicht eine Welt, sondern nur ein *Weltmodell* erleben, würde global verfügbar: Sie könnten diese Information deshalb jetzt auch zur Kontrolle Ihrer Handlungen und Ihrer Gedanken oder für die Aufmerksamkeitslenkung nutzen. Sie würden die *globale* Transparenz verlieren. Interessanterweise ist kognitive Verfügbarkeit allein aber nicht hinreichend, um den naiven Realismus des phänomenalen Erlebens zu durchbrechen. Man kann sich nicht einfach aus dem phänomenalen Modell der Wirklichkeit »herausdenken«, indem man seine Meinungen über dieses Modell ändert: Die Transparenz phänomenaler Repräsentationen ist kognitiv nicht penetraibel, phänomenales Wissen ist nicht dasselbe wie begrifflich-propositionales Wissen.

Diesen Gedanken muß man nun im letzten Schritt wieder auf das Selbstmodell anwenden. Meine Hauptthese ist: Wir sind Systeme, die erlebnismäßig nicht in der Lage sind, ihr eigenes subsymbolisches Selbstmodell *als* Selbstmodell zu erkennen. Aus diesem Grund operieren wir, phänomenologisch gesehen, unter den Bedingungen eines »naiv-realistischen Selbstmißverständnisses«: Wir erleben uns selbst, als wären wir in direktem und unmittelbarem epistemischen Kontakt zu uns selbst. Denn — das ist der begriffliche Punkt — ein phänomenal transparentes Selbstmodell erzeugt mit logischer Notwendigkeit das bewußte Erlebnis, *sich selbst unendlich nah zu sein*. Die Kernaussage der SMT ist, daß auf diese Weise ein basales »Ich-Gefühl«<sup>4</sup>. Entsteht,

---

<sup>4</sup>Auf begrifflicher und empirischer Ebene die konstitutiven bzw. (relativ zu einer spezifischen Klasse von Systemen) minimalen Bedingungen für die Entstehung eines phänomenalen Selbst zu isolieren, ist eines der

ein für das betreffende System unhintergebares phänomenales Selbst. Der Inhalt des nicht-begrifflichen Selbstbewußtseins ist der Inhalt eines transparenten PSM. Dies legt mich auch auf eine spezifische empirische Voraussage fest: Wenn das PSM seine Transparenz verlöre und opak würde, wenn der Organismus als Ganzer in der Lage wäre, sein momentanes Selbstmodell erlebnismäßig *als* ein Modell zu erkennen, dann würde die Eigenschaft der Selbstheit verschwinden, das »Ich-Gefühl« würde sich auflösen. In phänomenologischen Standardkonfigurationen aber ist die Entität, die auf das Buch in ihren Händen schaut, selbst eine Form von transparentem phänomenalem Inhalt. Und dies trifft ebenso auf die »Gerichtetheit«, die mit diesem Akt der visuellen Aufmerksamkeit einhergeht, zu — also auf die scheinbar direkte Beziehung, die Subjekt und Objekt zu verbinden scheint.

### Der sechste Schritt: Das phänomenale Modell der Intentionalitätsrelation

Gehen wir noch einen Schritt weiter, bevor wir zum Schluß kommen. Das phänomenale Erleben des »Ein-Selbst-Seins« ist aufs Engste nicht nur mit dem Gefühl des Besitzens verbunden, sondern auch mit dem Erleben von Agentivität. Es geht nicht nur darum, ein transparentes Selbstmodell zu besitzen, sondern auch um die Repräsentation von Gerichtetheit, d.h. darum, mit Zielobjekten und Zielzuständen dynamisch verbunden zu sein. Dazu nun zwei weitere empirische Beispiele, und zwar aus den experimentellen Neurowissenschaften, die Makaken als Versuchsobjekte benutzen.

In der klassischen Neurologie wurde die Hypothese eines »Körperschemas« aufgestellt, einer unbewußten, aber kontinuierlich aktualisierten Karte der Körperform und -haltung im Gehirn<sup>5</sup>. Neuere Stu-

---

wichtigsten Erkenntnisziele für die Selbstmodell-Theorie der Subjektivität. Dieses Forschungsprogramm ist mit dem Begriff der *minimal phenomenal selfhood* (MPS; Blanke & Metzinger, 2009) verknüpft. Ein guter Kandidat ist die phänomenal transparente Selbstlokalisierung in einem räumlichen und zeitlichen Bezugsrahmen, wobei Agentivität und das Vorhandensein eines expliziten Körpermodells keine notwendigen Bedingungen zu sein scheinen (vgl. Metzinger, 2013; Windt, 2010, 2014).

<sup>5</sup>Die Terminologie war nie ganz klar, aber sie unterscheidet oft zwischen einem unbewußten »Körperschema« und einem bewußten »Körperbild«. Eine philosophische Sicht auf die begriffliche Unklarheit, die beide Begriffe umgibt, findet sich in Gallagher (2005); einen guten Einstieg in die empirische Literatur bietet Maravita (2006). Eine ausgezeichnete neuere Überblicksdarstellung ist Iriki und Sakura (2008).

dien zeigen, wie japanische Makaken darauf trainiert werden können, Werkzeuge zu benutzen, obwohl sie in ihrem natürlichen Umfeld kaum Gebrauch von Werkzeugen machen (vgl. Maravita und Iriki, 2004). Während des erfolgreichen Gebrauchs von Werkzeugen finden Veränderungen in spezifischen neuronalen Netzen in ihren Gehirnen statt. Aus dieser Erkenntnis geht hervor, daß die Werkzeuge vorübergehend in ihr Körperschema integriert werden. Wenn ein Futterkugelchen außerhalb der Reichweite ihrer Hände ausgegeben wird und sie erfolgreich eine Harke benutzen, um es näher heranzuziehen, kann man eine Veränderung ihres körperlichen Selbstmodells im Gehirn beobachten. Es sieht tatsächlich so aus, als sei das bewußte Modell ihrer Hand bis zur Spitze des Werkzeugs erweitert worden. Eine genauere Art zu beschreiben, was hier geschieht, ist die folgende: Auf der Ebene des bewußten Realitätsmodells des Affen werden Eigenschaften der Hand auf die weiter entfernte Spitze des Werkzeugs übertragen. Wir kennen denselben Effekt bei Menschen: In unserem Fall kann wiederholte Übung die Spitze eines Werkzeuges zu einem Teil unserer eigenen Hand machen, zu einem Teil, der genauso »führend« und geschickt wie unsere eigenen Finger benutzt werden kann.

Mit anderen Worten: Neuere neurowissenschaftliche Daten stützen eindeutig die Ansicht, daß wir durch Werkzeuge nicht nur in der Lage sind, unsere Reichweite zu erweitern. Sie zeigen, daß jede erfolgreiche Erweiterung des Handlungsraums auch im neuronalen Träger des Körperbildes im Gehirn wiedergespiegelt wird. Das Gehirn konstruiert ein »internalisiertes« Bild des Werkzeugs, indem es rasch in das bereits existierende Körperbild als Ganzes integriert wird — es wird funktional »einverleibt«. Natürlich können wir nicht wirklich sagen, ob Affen tatsächlich bewußt das Gefühl des Besitzens erleben oder ob es sich hier nur um unbewußte Mechanismen handelt. Wir wissen aber um mehrere Ähnlichkeiten zwischen Makaken und Menschen, die diese Annahme plausibel machen. Dies könnte auch einer der Anfänge der mentalen *Simulation* einer momentanen Gerichtetheit auf ein Zielobjekt oder einen Zielzustand sein. Und dieser Schritt führt uns zum zweiten Hauptaspekt der Selbstheit oder des »Ich-Gefühls«: Neben globalem *Besitzen* müssen wir *Agentivität* verstehen, das Phänomen der flexiblen globalen Handlungskontrolle.

Ein spannender Aspekt dieser neuen Daten ist, daß sie Aufschluß geben über die Evolution des Werkzeuggebrauchs. Eine notwendige Voraussetzung für die Erweiterung des eigenen Handlungsraums und der eigenen Befähigungen durch Werkzeuggebrauch scheint nämlich ganz eindeutig die Fähigkeit zu sein, Werkzeuge in ein bereits gegebenes Selbstmodell zu integrieren. Man kann nur dann zielgerichtet und intel-

ligent Gebrauch von Werkzeugen machen, wenn das Gehirn sie zeitweise als Teil des eigenen Selbst repräsentiert. Intelligenter Werkzeuggebrauch war eine entscheidende Errungenschaft in der menschlichen Evolution. Man kann annehmen, daß eine Art elementarer funktionaler Baustein des menschlichen Werkzeuggebrauchs schon in den Gehirnen unserer Vorfahren existiert hat. Durch evolutionären Druck und noch nicht vollständig verstandene kulturelle Mechanismen hat sich dieser Grundbaustein dann schnell zu dem entwickelt, was wir heute bei Menschen beobachten<sup>6</sup>. Ich denke, einer der interessantesten Aspekte der neueren soziokulturellen Entwicklung besteht genau darin, daß Menschen sich mit Hilfe ihres neuartigen Selbstmodells in einem bis dahin ungekannten Ausmaß *gegenseitig* zu instrumentalisieren versuchen: Menschen versuchen nämlich sehr häufig, ihren eigenen Handlungsraum zu vergrößern, indem sie *andere* Menschen kontrollieren und als Werkzeuge einsetzen. Meine spekulative These ist, daß diese Tatsache die Evolution von Großgesellschaften auf entscheidende Weise angetrieben hat.

Es gibt heute ein schnell wachsendes Forschungsfeld, in dem Ingenieure und Neurowissenschaftler zusammenarbeiten: *Brain-Machine Interfaces* (ein kurzer Überblick findet sich in Lebedev & Nicolelis, 2006). Eine weitere Anwendung des oben skizzierten Grundgedankens besteht darin, künstliche Glieder oder Roboter direkt durch bestimmte Verbände von kortikalen Neuronen anzusteuern und zu kontrollieren, also eine Maschine in die Lage zu versetzen, motorische Befehle auszuführen, die in einem biologischen Gehirn erzeugt werden. Mittlerweile können Affen bereits über das Internet in Echtzeit die Schrittmuster von humanoiden Robotern auf der anderen Seite der Welt fernsteuern, von der *Duke University* in Amerika zum *Computational Brain Project* bei der *Japan Science and Technology Agency* in Japan, und zwar ausschließlich durch die Aufzeichnung ihrer Gehirnaktivität. Miguel Nicolelis schreibt:

Die verblüffendste Entdeckung war, daß der Affe, als wir das Laufband abschalteten und er aufhörte, seine Beine zu bewegen, in der Lage war, die Fortbewegung des Roboters für einige Minuten aufrechtzuerhalten — nur durch seine Gedanken und ausschließlich mit Hilfe des visuellen Feedbacks vom Roboter in Japan.

In unserem Zusammenhang ist die vielleicht interessanteste Beobachtung in diesem Experiment (zu den Details vgl. Carmena et al., 2003), daß der Affe allmählich seinen eigenen Arm vernachlässigt, obwohl er immer noch Teil seines biologischen Körpers ist. Das heißt, daß in seiner

---

<sup>6</sup>Vgl. Iriki et al., 1996; Maravita and Iriki, 2004.

Selbstrepräsentation im Gehirn bestimmte Veränderungen abzulaufen scheinen, weil er jetzt versucht, das Feedback in einer neuen Art von motorischer Aufgabe und mit einem anderen Zielzustand zu kontrollieren. Dabei optimiert er ein neues Set von motorischen Parametern, indem er versucht, einen realen Roboterarm oder sogar einen virtuellen Arm zu kontrollieren, den er auf dem Bildschirm vor sich sieht. Die Abstimmung der Neuronen, ihre *tuning properties*, verändern sich also. Lebedev und Nicoletis (2006, S. 542) beschreiben diesen Effekt wie folgt:

Es ist bemerkenswert, daß die Gliedmaßen der Tiere, nachdem diese angefangen hatten den Aktuator durch ihre neuronalen Aktivitäten direkt zu kontrollieren, aufhörten, sich zu bewegen, während die Tiere weiterhin den Aktuator durch das Generieren geeigneter Modulationen ihrer kortikalen Neuronen kontrollierten. Die sparsamste Interpretation dieser Ergebnisse ist die, daß das Gehirn in der Lage war, eine graduelle Assimilation des Aktuators in die Karten, die den Körper repräsentierten, durchzuführen.

Aus der Perspektive der SMT ist die plausibelste Interpretation die, daß die Repräsentationen der alten Körperteile des Affen bestimmte funktionale Eigenschaften verlieren, sobald das Tier eine innere Repräsentation dieses neuen Aktuators erfolgreich in sein bewußtes Selbstmodell einbettet. Die alten Körperteile werden vorübergehend immer weniger zugänglich für die Aufmerksamkeitsverarbeitung und treten nach und nach aus dem bewußten Erleben zurück. Diese Beispiele führen uns zwei weitere wichtige Einsichten vor Augen. Zum einen ist es offensichtlich, daß das phänomenale Selbstmodell oft ein wichtiger Teil einer *Kontrollhierarchie* ist: Es ist ein abstraktes Werkzeug, ein Mittel zur Überwachung bestimmter kritischer Aspekte desjenigen Prozesses, in dem der Organismus flexible, adaptive Verhaltensmuster erzeugt. Zum anderen ist es hochgradig plastisch, weil mehrere Repräsentationen von Objekten *außerhalb* des Körpers vorübergehend in das Selbstmodell integriert werden können. Dies trifft nicht nur auf Gummihände zu, sondern ist umso mehr auf Werkzeuge im Allgemeinen anwendbar — Erweiterungen körperlicher Organe, die erfolgreich kontrolliert werden müssen, um intelligentes und zielgerichtetes Verhalten zu erzeugen. Das Selbstmodell ist das funktionale Fenster, durch das das Gehirn mit dem Körper *als einer Ganzheit* interagieren kann, und umgekehrt. Wenn der Körper durch Stöcke, Steine, Harken oder Roboterarme erweitert wird, muß auch das Selbstmodell erweitert werden. Erst wenn eine integrierte Repräsentation des Körpers-plus-Werkzeug existiert, kann das erweiterte System des Körpers-plus-Werkzeug ein Teil der Kontrollhierarchie des Gehirns werden. Anders gefragt: Wie sollte man überhaupt lernen, ein Werkzeug intelligent — also: auf flexible und

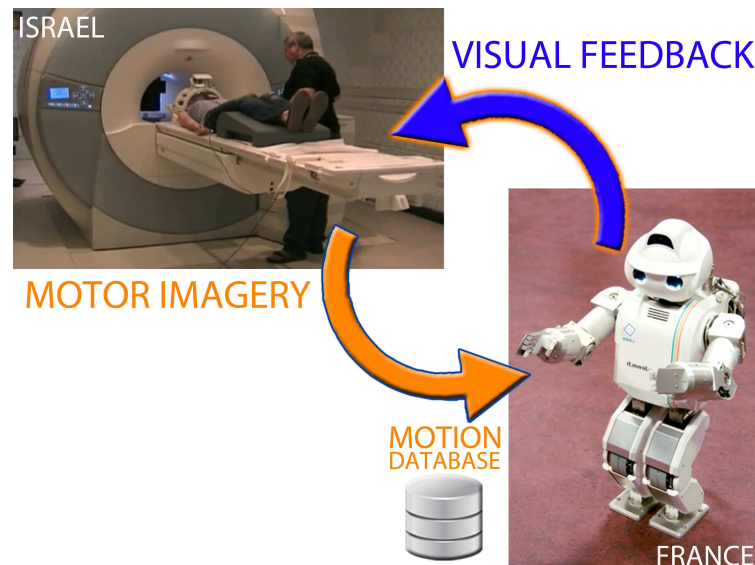
kontextsensitive Weise — zu benutzen, *ohne* es in das bewußte Selbst zu integrieren? Das bewußte Selbstmodell ist ein virtuelles Organ, das uns erlaubt, Feedbackschleifen zu *besitzen* und Kontrollprozesse zu initiieren, aufrechtzuerhalten und flexibel anzupassen. Einige Elemente des Kontrollkreislaufs sind physisch (wie das Gehirn und Werkzeuge), andere sind virtuell (wie das Selbstmodell und die Zielzustandssimulation).

Natürlich ist es unter SMT denkbar, auch komplette Roboter oder virtuelle Körper (Avatare) in das PSM einzubetten und dadurch kausal zu kontrollieren. Eine empirische Vorhersage unter den begrifflichen Annahmen von SMT ist nämlich, daß es prinzipiell möglich sein muß, das menschliche Selbstmodell unter Umgehung des nicht-neuronalen, biologischen Körpers in einer kausal direkteren Weise an künstliche Effektoren und Sensoren zu koppeln und dadurch funktional in einer technologisch erzeugten Umwelt zu situieren. In einer ehrgeizigen Pilotstudie haben Ori Cohen, Doron Friedman und ihre Kollegen demonstriert, daß es mit funktionaler Echtzeit-Magnetresonanztomographie möglich ist, die motorischen Intentionen einer Versuchsperson auszulesen und direkt als hochstufige Motorbefehle an einen humanoiden Roboter zu übertragen, der sie in körperliche Handlungen transformiert, während die Versuchsperson gleichzeitig das gesamte Experiment durch die Augen des Roboters visuell miterleben kann (Cohen et al., 2012). Der Vorgang basiert auf bildhaften Bewegungsvorstellungen, die es den Versuchspersonen erlauben, »direkt mit Ihrem PSM zu handeln«,<sup>7</sup> indem sie aus einem Scanner in Israel einen humanoiden HO-AP3-Roboter in Frankreich kontrollieren.

Oben habe ich bereits drauf hingewiesen, daß Menschen (und auch manche andere Tiere) oft das Verhalten oder die Befindlichkeit einer *anderen* Person kontrollieren wollen. Wir »instrumentalisieren« und »beschlagnahmen« einander, manchmal machen wir uns sogar zu »Leib-eigenen«. Menschen erweitern sich ständig — nicht nur mit Stöcken, Steinen, Harken oder Roboterarmen, sondern auch mit den Gehirnen

---

<sup>7</sup>An andere Stelle habe ich die ethischen Implikationen dieser historisch neuartigen Handlungsformen untersucht und den Begriff einer »PSM-action« eingeführt, einer Handlung, bei der ein menschlicher Akteur Teile seines Selbstmodells zu *offline*-Simulationen benutzt, die dann unter Umgehung des proximalen, nicht-neuronalen *embodiments* zu Konsequenzen in der Welt außerhalb des biologischen Körpers führt. Es ist aus verschiedenen Gründen denkbar, daß die Handlungsautonomie — und damit auch die ethische Verantwortlichkeit — des Akteurs im Sinne von Impulskontrolle und Abbruchmöglichkeiten in solchen Situationen eingeschränkt ist. Vgl. Metzinger, 2013a.



RYS. 1. Direktes Handeln mit dem PSM durch robotic re-embodiment: Das Ziel des Experiments bestand darin, es einer Versuchsperson in Israel durch »direkte Gedankenkontrolle« über das Internet zu ermöglichen, einen Roboter in Frankreich zu kontrollieren. Eine Video-Demonstration findet sich unter <http://www.youtube.com/user/TheAVL2011>. Abbildung mit freundlicher Genehmigung von Doron Friedman.

und Körpern *anderer* Menschen (Metzinger & Gallese, 2003). Der Übergang von biologischer zu kultureller Evolution ist ohne Frage unmittelbar mit der Realisierung neuer und spezifischer funktionaler Eigenschaften durch das PSM der Primaten verbunden. Dies ist eine der interessantesten Fragen für die Zukunft: Was genau war die Veränderung im PSM des *Homo sapiens*, gegenüber dem PSM des Schimpansen, die zur Explosion von Kultur und der Entwicklung komplexer Gesellschaften führte? Meine zweite spekulative Arbeitshypothese wäre hier die, daß es nicht allein komplexer Werkzeuggebrauch in der sozialen Dimension *per se* war, sondern auch die Fähigkeit, einen viel größeren Teil der Kontrollhierarchie *offline* laufen zu lassen, sie in Simulationen zu benutzen und gleichzeitig ein opakes (d.h. ein nicht-transparentes) PSM zu erzeugen<sup>8</sup>. Es war die Fähigkeit, sich selbst bewußt *als* re-

<sup>8</sup>Die SMT sagt, daß der Inhalt eines transparenten Zustandes als gegeben erlebt wird, der eines opaken aber als selbst konstruiert (Metzinger, 2003a, c; 2009). Ein interessanter Gedanke von Jennifer Windt beruht auf der Beobachtung, daß die frühen Werkzeuge unserer Vorfahren nicht nur benutzt, sondern — vor der Einbettung ins Selbstmodell — selbst erst hergestellt

präsentierend, als auf einen Zielzustand gerichtet zu repräsentieren. Es war der Unterschied zwischen dem bloßem Besitzen einer Erste-Person-Perspektive und der mentalen Fähigkeit, eben diese Tatsache noch einmal explizit zu repräsentieren.

Werfen wir deshalb einen Blick auf die repräsentationale Architektur, die dem subjektiven Erleben von Gerichtetheit im Allgemeinen zugrunde liegt. Phänomenologisch gesehen entsteht aus einem transparenten Weltmodell eine Wirklichkeit. Aus einem transparenten Modell des Systems entsteht ein Selbst, das in diese Wirklichkeit integriert ist. Wenn nun noch ein transparentes Modell der vorübergehenden und sich kontinuierlich verändernden Beziehungen zwischen dem wahrnehmenden und handelnden Selbst und den Objekten und Personen in dieser Wirklichkeit entsteht, resultiert dies in dem, was ich oben eine »phänomenale Erste-Person-Perspektive« genannt habe. Eine genuine Innenperspektive entsteht genau dann, wenn sich das System *für sich selbst* als momentan mit der Welt interagierend darstellt und wenn es diese Darstellung wiederum nicht als solche erkennt. Es besitzt dann ein bewußtes Modell der Intentionalitätsrelation<sup>9</sup>. Es stellt sich als auf bestimmte Aspekte der Welt gerichtet dar. Sein phänomenaler Raum ist ein *perspektivischer* Raum und seine Erlebnisse sind jetzt *subjektive* Erlebnisse.

Die Intentionalitätsrelation ist in der Hauptsache eine Wissensbeziehung zwischen Subjekt und Objekt: Ein mentaler Zustand wird dadurch zu einem Träger von Wissen, daß er über sich selbst hinaus verweist — wie ein Pfeil, der aus dem Geist eines Menschen auf ein Objekt in der realen oder sogar nur in einer möglichen Welt zeigt. Philosophen sagen, daß diese Art von mentalem Zustand einen *intentionalen Inhalt* besitzt. Der Inhalt ist das, worauf der Pfeil zeigt. Dieser Inhalt kann ein Bild, eine Aussage oder sogar ein Handlungsziel sein.

---

werden mußten. Affen dagegen stellen ihre eigenen Werkzeuge kaum oder nur sehr selten her. Die ersten Werkzeuge unserer Vorfahren wurden anscheinend in wenigen Schritten (4 bis 5 Steinschläge) hergestellt, komplexere Werkzeuge, die man dann für verschiedene Tätigkeiten nutzen und auch über längere Zeit behalten konnte (eine Voraussetzung für die geographische Ausbreitung), haben dann mehrere durchdachte Arbeitsschritte (und verschiedene andere Werkzeuge) erfordert — also wahrscheinlich auch mentale *offline*-Simulationen. *Trial and error* hätte wohl kaum zu dieser Komplexität und auch zu der Effizienz der Herstellung geführt, außerdem wurde bei der Benutzung dann tatsächlich etwas ins Selbstmodell eingebettet, das vorher durch eigenes körperliches Handeln und ganz konkret und sinnlich wahrnehmbar *selbst* konstruiert worden war.

<sup>9</sup>Vgl. dazu insbesondere den zweiten Teil von Metzinger 2006c.



Wie Philosophen sagen, gibt es »praktische Intentionalität« in dem Sinne, daß man auf bestimmte »Erfüllungsbedingungen« (wie z.B. ein Handlungsziel) gerichtet ist, und es gibt eine »theoretische Intentionalität« in dem Sinne, daß man auf »Wahrheitsbedingungen« (z.B. eines Satzes, also ein Erkenntnisziel) gerichtet ist. Wenn viele solcher Pfeile im Bewußtsein verfügbar sind, d.h. vom Gehirn auf der funktionalen Ebene der globalen Verfügbarkeit dargestellt werden, dann entsteht eine zeitlich ausgedehnte Erste-Person-Perspektive. Kurz gesagt: Es ist eine Sache, ein biologischer Organismus zu sein, der die Welt darstellt, es ist aber etwas ganz Anderes, sich selbst auf der Ebene des bewußten Erlebens als *darstellend* darzustellen, und zwar in Echtzeit und während der dynamische Vorgang der Repräsentation sich tatsächlich ereignet. Die SMT beschäftigt sich mit Letzterem. Es gibt dann nicht mehr nur ein neuronal verankertes Kernselbst, eine Selbstrepräsentation, sondern auch eine dynamische, phänomenale Simulation des *Selbst als eines Subjekts*, das durch ständig wechselnde Wissens- und Handlungsbeziehungen in die Welt eingebettet ist. Natürlich wäre zum zentralen Konzept eines »phänomenalen Modells der Intentionalitätsrelation« (PMIR) sehr viel mehr zu sagen<sup>10</sup>. Der Kerngedanke ist aber der folgende: Der bewußte Mensch ist ein System, das bei einzelnen repräsentationalen Akten die Repräsentationsbeziehung selbst noch einmal dynamisch *ko-repräsentieren* kann. Das Instrument, das es für diesen Zweck benutzt, ist das PMIR. Das PMIR ist lediglich ein weiteres virtuelles Organ, genau wie das PSM. Der Inhalt höherstufiger Formen des Selbstbewußtseins ist nämlich immer eine Relation: Das Selbst *im Moment des Erkennens* (Damasio, 1999, S. 168ff), das Selbst *im Akt des Handelns*. Ein Subjekt zu sein bedeutet, die Fähigkeit zu besitzen, diese intentionale Beziehung selbst zu ko-repräsentieren und sie gleichzeitig durch Interaktion mit einer Welt aktiv zu konstruieren.

Natürlich ist die Art und Weise, in der wir diese Subjekt-Objekt-Beziehungen subjektiv erleben, eine vereinfachte Version der eigentlichen Prozesse — sie ist gewissermaßen eine funktional adäquate Konfabulation. Die Evolution hat auch in diesem Fall wieder eine einfache, aber elegante Lösung favorisiert. Das virtuelle Selbst, das sich in der phänomenalen Welt bewegt, besitzt kein Gehirn, kein Motorsystem und keine Sinnesorgane: Bestimmte Teile der Umgebung erscheinen direkt

---

<sup>10</sup>Die PMIR-Theorie ist natürlich komplexer, als ich es in diesem kurzen Überblicksartikel darstellen kann. Neben Metzinger (2003a) empfehle ich Lesern, die sich für die Idee interessieren, Metzinger (2006c). Eine detailliertere Besprechung, die sich besonders auf die repräsentationale Struktur bewußter Willensakte bezieht, ist zu finden in Metzinger (2006a).

in seinem Geist, der Wahrnehmungsprozeß wird als anstrengungslos und unmittelbar erlebt. Auch Körperbewegungen werden scheinbar »direkt« ausgelöst. Solche Effekte sind typisch für *unsere* Form des subjektiven Erlebens und sie sind — als neurokomputationale Strategie betrachtet — die Vorteile einer benutzerfreundlichen Oberfläche. Was oben als »Transparenz« definiert wurde, ist eine Art, die *Geschlossenheit* dieser multimodalen, hochdimensionalen Benutzeroberfläche des Gehirns zu beschreiben. Das phänomenale Selbst ist der Teil dieser Oberfläche, den das System benutzt, um *sich selbst* als ein Ganzes zu repräsentieren und sein eigenes Verhalten vorherzusagen, um sich für sich selbst als ein erkennendes »Ich« darzustellen und sich selbst als Agenten zu begreifen. Dieser virtuelle Agent »sieht mit den Augen« und »handelt mit den Händen«. Er weiß nicht, daß er einen visuellen oder einen motorischen Kortex besitzt. Das PSM ist das interne Interface, welches das System benutzt, um sich seine eigene Hardware funktional anzueignen, um seine eigenen niedrigstufigen Dynamiken zu kontrollieren, um zeitliche entfernte Zielzustände als *seine eigenen* darzustellen und um *autonom* zu werden. Die intentionalen Pfeile, die diesen Agenten mit Objekten und anderen Selbsten innerhalb des gerade aktiven Wirklichkeitsmodells verbinden, sind phänomenale Repräsentationen von vorübergehend auftretenden Subjekt-Objekt-Beziehungen — und oft können auch sie nicht als Repräsentationsprozesse erkannt werden. In Standardsituationen ist die bewußt erlebte Erste-Person-Perspektive der Inhalt eines transparenten PMIR.

All dies spielt sich innerhalb eines phänomenalen Gegenwartsfensters ab. Die Inhalte des phänomenalen Erlebens erzeugen nicht nur eine Welt, sondern auch eine *Gegenwart* (vgl. Metzinger, 2003a, Abschnitt 3.2.2). Wenn man so will, ist der Kern des phänomenalen Bewußtseins die Erzeugung einer Gegenwartsinsel im physikalischen Fluß der Zeit (vgl. Ruhnau, 1995 und die dort gegebenen Literaturhinweise, insbesondere auf die Arbeiten von Ernst Pöppel). Minimales Selbstbewußtsein ist Selbstlokalisierung innerhalb eines räumlichen und eines zeitlichen Bezugsrahmens (Blanke & Metzinger, 2009): Subjektives Erleben heißt »hier sein« (Immersion in einer Szene) und immer auch »jetzt sein« (Präsenz). Bewußtsein ist Gegenwärtigsein. Es bedeutet, Informationen in einer sehr speziellen Weise zu verarbeiten. Diese besteht darin, bereits repräsentierte Einzelereignisse immer wieder und kontinuierlich zu zeitlichen Gestalten zusammenzufassen, zu einem bewußt erlebten Moment. Viele empirische Daten zeigen heute, daß die bewußt erlebte Gegenwart in einem bestimmten Sinne eine *erinnerte* Gegenwart ist. So gesehen ist sogar das phänomenale »Jetzt« ein repräsentationales Konstrukt, eine *virtuelle* Gegenwart. An diesem

Punkt wird erstmals deutlich, was es bedeutet, wenn man sagt, daß der phänomenale Raum ein virtueller Raum ist: Sein Inhalt ist eine *mögliche* Realität, eine Vorhersage<sup>11</sup>. Der Realismus des phänomenalen Erlebens entsteht dadurch, daß in ihm eine Möglichkeit — die beste Hypothese, die es im Moment gibt — unhintergebar als Wirklichkeit oder als eine *Aktualität* dargestellt wird. Mit anderen Worten: Die Mechanismen, die zeitliches Erleben und unser subjektives Gegenwartsgefühl herstellen, sind ebenfalls transparent. Auch diesen Punkt muß man nun am Ende auf den Sonderfall der phänomenalen Selbstmodellierung anwenden: Weil die Virtualität des Selbstmodells und die Virtualität des Gegenwartsfensters nicht auf der Ebene des subjektiven Erlebens selbst verfügbar sind, wird das in ihnen dargestellte System zu einem *jetzt anwesenden Subjekt*.

SMT löst das Homunkulus-Problem, weil wir jetzt sehen können, daß kein »kleiner Mann im Kopf« gebraucht wird, um den Inhalt mentaler Repräsentationen zu interpretieren und »auszulesen«. Gleichzeitig ist sie ein höchst sparsames Erklärungsmodell, da sie es uns erlaubt, die Entstehung von Selbstbewußtsein zu erklären, ohne anzunehmen, daß ein substantielles Selbst existiert und mit dem PSM gleichzeitig einen für zukünftige Entwicklungen offenen Nachfolgebegriff anbietet.<sup>12</sup> Soll all dies bedeuten, daß das Selbst – ontologisch gesehen – nur eine Illusion ist? Auf den zweiten Blick enthalten der allzu populäre Begriff der »Ich-Illusion« und die Metapher des »Sich- mit-seinem-eigenen-inneren-Bild-von-sich-selbst-Verwechselns« einen logischen Fehler: *Wessen* Illusion sollte dies sein? Von einer Illusion zu sprechen setzt die Annahme voraus, daß es jemanden gibt, der sie *hat*. Etwas, das kein epistemisches Subjekt in einem starken Sinne begrifflich-propositionalen Wissens ist, *kann* sich überhaupt nicht mit irgendetwas anderem verwechseln. Wahrheit und Falschheit, Realität und Illusion gibt es für ein biologisches Informationsverarbeitungssystem auf der fraglichen Entwicklungsstufe noch gar nicht. Bisher haben wir ja nur eine kognitionswissenschaftlich fundierte Theorie der Phänomenologie der Selbstheit, des »Ich-Gefühls«, aber noch keine Theorie der Selbsterkenntnis. Ich habe hier lediglich kurz ausgeführt, wie eine *phänomenale* Erste-Person-Perspektive das Produkt der natürlichen Evolution

---

<sup>11</sup>Meine eigenen Vorstellungen ähneln in diesem Punkt denen von Antti Revonsuo *Virtuelle Realität* ist die beste technische Metapher für phänomenales Bewußtsein, die wir gegenwärtig besitzen. Vgl. Revonsuo (1995, 2000), und besonders Revonsuo (2006).

<sup>12</sup>Zu den verschiedenen Weisen, auf die man ein Anti-Realist bezüglich »des« Selbst sein kann vgl. Metzinger, 2011.

sein könnte. Subjektivität im *epistemischen* Sinne, eine epistemische Erste-Person-Perspektive, wäre ein weiterer Schritt. Natürlich ist die Phänomenologie des »Ich-Gefühls«, des nicht-begrifflichen Selbstbewußtseins, die wichtigste Voraussetzung für diesen Schritt, weil es die zentrale notwendige Bedingung für die Konstitution eines genuin *reflexiven*, begrifflichen Selbstbewußtseins ist. In gewisser Hinsicht besteht darin der ganze Sinn der Theorie: Wenn wir die höheren Formen von Subjektivität und Intersubjektivität ernst nehmen wollen, müssen wir zunächst bescheiden und vorsichtig sein, uns auf die fundierenden Ursprünge auf der Ebene nicht-begrifflicher Inhalte und der selbstorganisierenden neuronalen Dynamik im Gehirn konzentrieren. Meine Leser werden nicht überrascht sein, wenn der Autor dieses Aufsatzes der Überzeugung ist, daß subjektives *Wissen* am Ende dann eben genau das Wissen ist, das mit einem spezifischen inneren Modus der Präsentation zusammenhängt, nämlich als Wissen *unter einem phänomenalen Modell der Intentionalitätsrelation*. Was wir heute die »Erste-Person-Perspektive« nennen, ist eine hochspezifische, neuronal realisierte Weise des Gegebenseins, ein abstraktes Datenformat, welches im Laufe der biologischen Evolution schrittweise entstanden ist. Subjektivität im erkenntnistheoretischen Sinne kann deshalb zumindest prinzipiell ebenfalls naturalisiert werden. Dies ist aber nur dann möglich, wenn wir eine überzeugende evolutionäre und neurowissenschaftliche Erklärung dafür finden, wie sich genau diese repräsentationale Struktur, der dynamische, indexikalische innere Modus des Gegebenseins, den wir heute »die subjektive Innen-perspektive« nennen, in einem selbstorganisierenden physikalischen Universum überhaupt erst entwickeln konnte. Worum es geht, ist, daß ein natürlich evolviertes System beginnt, sich intern selbst *als epistemischen Agenten* zu modellieren, und dabei gleichzeitig mit Hilfe des PSM und des PMIR eine neue Modalität des Erkennens erzeugt, eine neue Form der Darstellung von Wissen über die Welt. Letztlich ist es offensichtlich, daß jedes einzelne Vorkommnis eines PSM bzw. PMIR identisch ist mit einem spezifischen Zeitfenster in der kontinuierlichen, dynamischen Selbstorganisation kohärenter Aktivität, die lokal in einem einzelnen biologischen Gehirn stattfindet. In diesem kontinuierlich fortlaufenden Prozeß auf der subpersonalen Ebene gibt es aber keinen kleinen Agenten und auch keinen bösen Dämon, der als *Urheber* einer Illusion angesehen werden könnte. Und es gibt auch keine starre subpersonale Entität, die als das *Subjekt* der Illusion angesehen werden könnte. Es gibt niemanden *im* System, der sich täuschen oder mit etwas verwechseln könnte, der Homunkulus existiert nicht. Selbstbewußtsein ist eine Eigenschaft des Systems als Ganzem.

## Literatur

- Ainley, V., Tajadura-Jiménez, A., Fotopoulou, A. & Tsakiris, M. (2012). Looking into myself: Changes in interoceptive sensitivity during mirror self-observation. *Psychophysiology* 49(11): 1672–6.
- Alvarado, C.S. (2000). Out-of-Body Experiences, in E. Cardena et al. (eds.), *Varieties of Anomalous Experience: Examining the Scientific Evidence*. Washington, DC: American Psychological Association.
- Armel, K.C. & Ramachandran, V.S. (2003). Projecting sensations to external objects: Evidence from skin conductance response. *Proceedings of the Royal Society of London: Biological* 270: 1499–1506.
- Baars, B.J. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- Baker, L.R. (1998). The first-person perspective: A test for naturalism. *American Philosophical Quarterly* 35: 327–46.
- Baker, L.R. (2000). Die Perspektive der ersten Person: Ein Test für den Naturalismus. In G. Keil und H. Schnädelbach (Hrsg.), *Naturalismus — Philosophische Beiträge*. Frankfurt am Main: Suhrkamp.
- Baker, L.R. (2008). Naturalism and the first-person perspective. In G. Gasser (ed.), *How successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society*. Frankfurt am Main: Ontos Verlag.
- Bennett, M.R. & Hacker, P.S.M. (2010). *Die philosophischen Grundlagen der Neurowissenschaften*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Bermúdez, J.L., Marcel, A. & Eilan, N. (1995)[eds]. *The Body and the Self*. Cambridge, MA: MIT Press.
- Bieri, P. (1987). Evolution, Erkenntnis und Kognition. In W. Lütterfelds (Hrsg.), *Transzendente oder Evolutionäre Erkenntnistheorie?* Darmstadt: Wissenschaftliche Buchgesellschaft.
- Bischof-Köhler, D. (1989, 2<sup>1</sup>993). Spiegelbild und Empathie. Bern: Huber.
- Bischof-Köhler, D. (1996). Ichbewußtsein und Zeitvergegenwärtigung. Zur Phylogese spezifisch menschlicher Erkenntnisformen. In A. Barkhaus, M. Mayer, N. Roughley & D. Thürnaue (Hrsg.), *Identität, Leiblichkeit, Normativität. Neue Horizonte anthropologischen Denkens*. Frankfurt am Main: Suhrkamp.
- Blanke, O. (2012). Multisensory mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience* 13(8): 556–571.
- Bongard, J., Zykov, V. & Lipson, H. (2006). Resilient machines through continuous self-modeling. *Science* 314: 1118.
- Botvinick, M. & Cohen, J. (1998). Rubber hands »feel« touch that eyes see. *Nature* 391: 756.
- Botvinick, M. (2004). Probing the neural basis of body ownership. *Science* 305: 782–3.
- Brugger, P., Regard, M., and Landis, T. (1997). Illusory reduplication of one's own body: phenomenology and classification of autoscopic phenomena. *Cognitive Neuropsychiatry* 2: 19–38.

Brugger, P., Kollias, S.K., Müri, R.M., Crelier, G., Hepp-Reymond, M.-C., and Regard, M. (2000). Beyond re-membering: Phantoms sensations of congenitally absent limbs. *Proceedings of the National Academy of Science USA* 97: 6167–72.

Brugger, P., Regard, M., and Shiffrar, M. (2001). Hand movement observation in a person born without hands: Is body scheme innate? *Journal of Neurology, Neurosurgery, and Psychiatry* 70: 276.

Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., Patil, P. G., Henriquez, C. S., and Nicolelis, M. A. L. (2003). Learning to Control a Brain-Machine Interface for Reaching and Grasping by Primates. *PLoS Biology* 1:193–208.

Churchland, P.M. (1989). *A Neurocomputational Perspective*. Cambridge, MA / London: MIT Press.

Clark, A. (1989). *Microcognition — Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge, MA: MIT Press.

Cohen, O., Druon, S., Lengagne, S., Mendelsohn, A., Malach, R., Kheddar, A. & Friedman, D. (2012). MRI-based robotic embodiment: A pilot study. *IEEE International Conference on Biomedical Robotics and Biomechanics*, June 24–27, Roma, Italy.

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press.

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

Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt Brace & Company.

Dennett, D.C. (1987). *The Intentional Stance*. Cambridge, MA und London: MIT Press.

Devinsky, O., Feldmann, E., Burrowes, K., and Bromfield, E. (1989). Autosopic phenomena with seizures. *Archives of Neurology* 46: 1080–8.

Dittrich, A. (1985). *Ätiologie-unabhängige Strukturen veränderter Wachbewußtseinszustände*. Stuttgart: Enke.

Dittrich, A. (1996). *Ätiologie-unabhängige Strukturen veränderter Wachbewußtseinszustände. Ergebnisse empirischer Untersuchungen über Halluzinogene I. und II. Ordnung, sensorische Deprivation, hypnagoge Zustände, hypnotische Verfahren sowie Reizüberflutung*. Berlin: VWB.

Dittrich, A., Lamparter, D., Maurer & M. (2006). 5D-ABZ: *Fragebogen zur Erfassung Außergewöhnlicher Bewußtseinszustände*. Zürich: PSIN PLUS.

Dretske, F. (1988). *Explaining Behavior — Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Dretske, F. (1998). *Die Naturalisierung des Geistes*. Paderborn: mentis.

Edelman, G.M. (1989). *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.

Ehrsson, H.H., Spence, C. & Passingham, R.E. (2004). That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science* 305: 875–7.

- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience* 11: 127–38.
- Gallagher, S. (2005). *How the Body shapes the Mind*. Oxford: Oxford University Press.
- Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences*, 2, 493–501.
- Gallese, V. (2005). Embodied simulation: from neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences* 4: 23–38.
- Gallup, G.G., Anderson, J.R., Platek, S.M. (2011). Self-recognition. In S. Gallagher (ed.), *The Oxford Handbook of the Self*. Oxford: Oxford University Press.
- Gallup, G.G., Jr. (1997). On the rise and fall of self-conception in primates. In J.G. Snodgrass and R.L. Thompson (eds.), *The Self across Psychology — Self-Recognition, Self-Awareness, and the Self Concept*. New York, NY: Annals of the New York Academy of Sciences 818: 4–17.
- Grush, R. (1997). The architecture of representation. *Philosophical Psychology* 10: 5–25.
- Grush, R. (1998). Wahrnehmung, Vorstellung, und die sensomotorische Schleife. In H.-D. Heckmann F. & Esken (Hrsg.), *Bewußtsein und Repräsentation*. Paderborn: mentis.
- Guimera, M.J., et al. (2011). Body integrity identity disorder: deranged body processing, right fronto-parietal dysfunction, and phenomenological experience of body incongruity. *Neuropsychological Review* 21: 320–333.
- Hilti, L. M., Hanggi, J., Vitacco, D. A., Kraemer, B., Palla, A., Luechinger, R., . . . (2012). The desire for healthy limb amputation: structural brain correlates and clinical features of xenomelia. *Brain* 136(1): 318–329.
- Howhy, J. (2013). *Mind in Prediction*. Oxford: Oxford University Press.
- Iriki, A., Tanaka, M., and Iwamura, Y. (1996). Coding of modified body schema during tool-use by macaque post-central neurons. *Neuroreport* 7: 2325–2330.
- Iriki, A. & Sakura, O. (2008). The neuroscience of primate intellectual evolution: natural selection and passive and intentional niche construction, *Phil. Trans. R. Soc. B* 363: 2229–2241.
- Jaspers, K. (1946). *Allgemeine Psychopathologie*. Berlin und Heidelberg.
- Kinsbourne, M. (1995). Awareness of one's own body: An attentional theory of its nature, development, and brain basis. In Bermúdez et al. 1995.
- Lenggenhager, B., Tadi, T., Metzinger, T. & Blanke, O. (2007). Video Ergo Sum: Manipulating bodily self-consciousness. *Science* 317: 1096–1099.
- Lycan, W.G. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.
- Maravita, A., and Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences* 8: 79–86.
- Maravita, A. (2006). From »body in the brain« to »body in space«: Sensory and intentional components of body representation. In G. Knoblich, I. Thornton, M. Grosjean & M. Shiffrar (eds.), *Body Perception from the Inside Out*. New York: Oxford University Press.

- Melzack, R. (1989). Phantom limbs, the self and the brain: The D.O. Hebb memorial lecture. *Canadian Psychology* 30: 1–16.
- Melzack, R. (1992). Phantom limbs. *Scientific American* 266: 90–6.
- Melzack, R., Israel, R., Lacroix, R. & Schultz, G. (1997). Phantom limbs in people with congenital limb deficiency or amputation in early childhood. *Brain* 120 (Pt 9): 1603–20.
- Melzack R. (2005). Evolution of the neuromatrix theory of pain. The Privithi Raj Lecture: presented at the Third World Congress of World Institute of Pain, Barcelona 2004. *Pain Pract* 5: 85–94.
- Metzinger, T. (1993; 2<sup>1</sup>999). *Subjekt und Selbstmodell. Die Perspektivität phänomenalen Bewußtseins vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation*. Paderborn: mentis.
- Metzinger, T. (1995a). *Perspektivische Fakten? Die Naturalisierung des »Blick von nirgendwo«*. In G. Meggle & J. Nida-Rümelin (1997)[Hrsg.], *ANALYOMEN 2 — Perspektiven der Analytischen Philosophie*. Berlin und New York: de Gruyter. S. 103–10.
- Metzinger, T. (1995b)[Hrsg.]. *Bewußtsein — Beiträge aus der Gegenwart-sphilosophie*. Paderborn: mentis.
- Metzinger, T. (1996). Niemand sein. In S. Krämer (Hrsg.), *Bewußtsein — Philosophische Positionen*. Frankfurt am Main: Suhrkamp.
- Metzinger, T. (1997). Ich-Störungen als pathologische Formen mentaler Selbstmodellierung. In Northoff, G. [Hrsg.], *Neuropsychiatrie und Neurophilosophie*. Paderborn: mentis.
- Metzinger, T. (2000a). The *subjectivity* of subjective experience: A representationalist analysis of the first-person perspective. In T. Metzinger (ed.), *Neural Correlates of Consciousness — Empirical and Conceptual Questions*. Cambridge, MA: MIT Press. Überarbeitete Version (2004): *Networks* 3–4: 33–64.
- Metzinger, T. (2000a). Die Selbstmodell-Theorie der Subjektivität: Eine Kurzdarstellung für Nicht-Philosophen in fünf Schritten. In W. Greve (Hrsg.), *Psychologie des Selbst*. Weinheim: BELTZ / Psychologie Verlags Union.
- Metzinger, T. (2000b) (ed.). *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.
- Metzinger, T. (2003). Phänomenale Transparenz und kognitive Selbstbezugnahme. In U. Haas-Spohn (Hrsg.), *Intentionalität zwischen Subjektivität und Weltbezug*. Paderborn: mentis.
- Metzinger, T. & Gallese, V. (2003). The emergence of a shared action ontology: building blocks for a theory. In G. Knoblich, B. Elsner, G. von Aschersleben & T. Metzinger (eds.), *Self and Action*. Special issue of *Consciousness & Cognition* (12:4), December 2003, 549–571.
- Metzinger, T. (2003a; 2<sup>2</sup>004). *Being No One. The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.
- Metzinger, T. (2003b). Phänomenale Transparenz und kognitive Selbstbezugnahme. In U. Haas-Spohn (Hrsg.), *Intentionalität zwischen Subjektivität und Weltbezug*. Paderborn: mentis. Seite 411–459.



Metzinger, T. (2003c). Phenomenal transparency and cognitive self-reference. *Phenomenology and the Cognitive Sciences* 2: 353–393.

Metzinger, T. (2004a). Why are identity-disorders interesting for philosophers? In T. Schramme und J. Thome (eds.), *Philosophy and Psychiatry*. Berlin: de Gruyter.

Metzinger, T. (2004b). Appearance is not knowledge: The incoherent strawman, content-content confusions and mindless conscious subjects. Invited commentary for Alva Noë und Evan Thompson: »Are there neural correlates of consciousness?«, in a special issue of *Journal of Consciousness Studies* 11(1): 67–71.

Metzinger, T. (2005b). Out-of-body experiences as the origin of the concept of a »soul«. *Mind and Matter* 3(1): 57–84.

Metzinger, T. (2005c). Die Selbstmodell-Theorie der Subjektivität: Eine Kurzdarstellung in sechs Schritten. In C. S. Herrmann, M. Pauen, J. W. Rieger & S. Schicktanz (Hrsg.), *Bewußtsein: Philosophie, Neurowissenschaften, Ethik*. Stuttgart: UTB/Fink .

Metzinger, T. (2006a). Conscious volition and mental representation: Towards a more fine-grained analysis. In N. Sebanz und W. Prinz (ed.), *Disorders of Volition*. Cambridge, MA: MIT Press.

Metzinger, T. (2006b). Reply to Gallagher: Different conceptions of embodiment. In *PSYCHE — An Interdisciplinary Journal of Research on Consciousness*, 12 (4).

Metzinger, T. (2006c). Being No One — Eine sehr kurze deutsche Zusammenfassung. In T. Metzinger, *Grundkurs Philosophie des Geistes, Band 1: Phänomenales Bewusstsein*. S. 424–475.

Metzinger, T. (2008). Empirical perspectives from the self-model theory of subjectivity: A brief summary with examples. In R. Banerjee & B.K. Chakrabarti (eds.), *Progress in Brain Research* 168: 215–246. Amsterdam: Elsevier.

Metzinger, T. (2009). *Der Ego Tunnel. Eine neue Philosophie des Selbst: Von der Hirnforschung zur Bewußtseinsethik*. Berlin: Bloomsbury Verlag.

Metzinger, T. (2011). *The No-Self-Alternative* (Chapter 11). In S. Gallagher (ed.), *The Oxford Handbook of the Self*. Oxford, UK: Oxford University Press. S. 279–296.

Metzinger, T. (2013a). Two principles for robot ethics. In E. Hilgendorf & J.-P. Günther (Hrsg.), *Robotik und Gesetzgebung*. Baden-Baden: Nomos.

Metzinger, T. (2013b). First-order embodiment, second-order embodiment, third-order embodiment: From spatiotemporal self-location to minimal phenomenal selfhood. In Lawrence Shapiro (ed.), *The Routledge Handbook of Embodied Cognition*. London: Routledge.

Michal, M. & Beutel, M.E. (2009). Weiterbildung CME: Depersonalisation/Derealisation — Krankheitsbild, Diagnostik und Therapie. *Z Psychosom Med Psychother* 55: 113–140.

Millikan, R.G. (1984). *Language, Thought, and other Biological Categories*. Cambridge MA: MIT Press.

- Millikan, R.G. (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.
- Nagel, T. (1986). *Der Blick von nirgendwo*. Frankfurt am Main: Suhrkamp.
- Newen, A. & Vogeley, K. (2000)[Hrsg.]. *Das Selbst und seine neurobiologischen Grundlagen*. Paderborn: mentis.
- Noreika, V., Windt, J.M., Lenggenhager, B., & Karim, A.A. (2010). New perspectives for the study of lucid dreaming: from brain stimulation to philosophical theories of self-consciousness. Commentary on »The neurobiology of consciousness: Lucid dreaming wakes up« by J. Allan Hobson. *International Journal of Dream Research* 3(1) 36–46.
- O’Shaughnessy, B. (1995). Proprioception and the body image. In Bermúdez *et al.* 1995.
- Ramachandran, V.S. & Blakeslee, S. (1998). *Phantoms in the Brain*. New York: William Morrow and Company, Inc.
- Ramachandran, V.S. & Rogers-Ramachandran, D. (1996). Synaesthesia in phantom limbs induced with mirrors. *Proceedings of the Royal Society London B*: 377–86.
- Ratcliffe, M. (2008). *Feelings of Being: Phenomenology, Psychiatry and the Sense of Reality*. Oxford & New York: Oxford University Press.
- Revonsuo, A. (1995). Consciousness, dreams, and virtual realities. *Philosophical Psychology* 8: 35–58.
- Revonsuo, A. (2000). Prospects for a scientific research program on consciousness. In T. Metzinger (ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.
- Revonsuo, A. (2006). *Inner Presence*. Cambridge, MA: MIT Press.
- Ruhnau, E. (1995). *Time-Gestalt and the observer*. In Metzinger 1995.
- Schilling, M. & Cruse, H. (2012). What’s next: recruitment of a grounded predictive body model for planning a robot’s actions. *Frontiers in Psychology* 3: 383. doi: 10.3389/fpsyg.2012.00383
- Seth, A.K., Suzuki, K. & Critchley, H.D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology* 2(395). doi: 10.3389/fpsyg.2011.00395
- Simeon, D., & Abugel, J. (2006). *Feeling Unreal: Depersonalization Disorder and the Loss of the Self*. Oxford/New York: Oxford University Press.
- Steels, L., & Hild, M. (2012). *Language Grounding in Robots*. Boston, MA: Springer US.
- Studerus, E., Gamma, A. & Vollenweider, F.X. (2010). Psychometric evaluation of the altered states of consciousness rating scale (OAV). *PLoS ONE* 5(8): e12412.
- Trivers, R. (2012). *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life*. New York: Basic Books. Deutsch: *Betrug und Selbstbetrug: Wie wir uns selbst und andere erfolgreich belügen* (2013; Berlin: Ullstein).
- Tsakiris, M. (2010). My body in the brain: a neurocognitive model of body-ownership. *Neuropsychologia* 48(3): 703–12.

Tsakiris, M. (2011). The sense of body ownership. In S. Gallagher (ed.), *The Oxford Handbook of the Self*. Oxford: Oxford University Press.

Tsakiris, M., Jiménez, A.T. & Costantini, M. (2011). Just a heart-beat away from one's body: interoceptive sensitivity predicts malleability of body-representations. *Proceedings of the Royal Society B: Biological Sciences* 278(1717): 2470–2476.

Tsakiris, M., Longo, M.R. & Haggard, P. (2010). Having a body versus moving your body: neural signatures of agency and body-ownership. *Neuropsychologia* 48(9): 2740–9.

Vallar, G. & Ronchi, R. (2009). Somatoparaphrenia: a body delusion. A review of the neuropsychological literature. *Experimental Brain Research* 192(3): 533–551.

Von Hippel, B., Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences* 34: 1–56.

Vuilleumier, P., Despland, P.A., Assal, G., and Regli, F. (1997). Héautoscopie, exta-se et hallucinations expérientielles d'origine épileptique. *Revue Neurologique* 153: 115–119.

Waelti, E. (1983) *Der dritte Kreis des Wissens*. Interlaken: Ansata.

Windt, J.M. (2010). The immersive spatiotemporal hallucination model of dreaming. *Phenomenology and the Cognitive Sciences* 9: 295–316.

Windt, J.M. & Metzinger, T. (2007). The philosophy of dreaming and self-consciousness: What happens to the experiential subject during the dream state? In D. Barrett & P. McNamara (eds.), *The New Science of Dreaming*. Estport, CT: Praeger Imprint/Greenwood Publishers.

Windt, J.M. (2014). *Dreaming*. Cambridge, MA: MIT Press.

Yates, J. (1985). The content of awareness is a model of the world. *Psychological Review* 92: 249–84.

## Empirical perspectives from the self-model theory of subjectivity: a brief summary with examples

THOMAS METZINGER  
*Johannes Gutenberg-Universität Mainz*

**Abstract.** *The fourth step consists in showing how the human self-model is a grounded model, because it is functionally anchored in low-level body-representation and interoceptive mechanisms like vestibular, proprioceptive, somatovisceral and homeostatic self-perception. The fifth step then introduces the most important, additional constraint, the theory of “phenomenal transparency” (see Metzinger 2003a: 3.2.7 and Metzinger 2003b: 354pp for details). “Transparency” refers to a property of conscious representations, namely, that they are not experienced as representations. Therefore, the subject of experience feels as if being in direct and immediate contact with their content. Transparent conscious representations create the phenomenology of naïve realism. An opaque phenomenal representation is one that is experienced as a representation, for example in pseudo-hallucinations, synesthetic concurrents, or lucid dreams, and — most importantly — in “conscious thoughts”, phenomenally experienced mental representations. It is the accessibility of earlier processing stages for attentional processing that leads to the phenomenal quality of “unrealness” or “mind-dependence”. Unconscious representations are neither transparent nor opaque. Moreover, there exists a graded spectrum between transparency and opacity, determining the variable phenomenology of “mind-independence” or “realness”. This phenomenology of “realness” can be analysed as the fact that the system has no introspective access to non-intentional properties of its own representations, that it is necessarily unaware of the construction process. If we now move from the representationalist to the functionalist level of analysis, we may interpret experiential “realness” as the Bayes-optimality, counterfactual richness, or plain dynamical stability of the generative models employed, or perhaps also as the functional property that further optimization of precision expectations is not possible. My own claim is that transparency is the phenomenal signature of epistemic reliability, and that if applied to the self-model it explains the phenomenology of “selfhood” and “realness” on the level of global self-representation.*

*The sixth and last step introduces the second major theoretical entity, the “phenomenal model of the intentionality relation” (PMIR).*

*Phenomenologically, a transparent world-model gives rise to a reality. A transparent system-model gives rise to a phenomenal self, embedded in this reality. If there is also a transparent model of the transient and constantly changing relations between the perceiving and acting self and the objects and persons in this reality, this results in the conscious experience of being not only a self, but also a subject, in the appearance of a “phenomenal first-person perspective”. A genuine inner perspective arises if only and only if the system represents itself as currently interacting with the world to itself, and if it does not recognize this representation as a representation. Now, it has a conscious model of the intentionality relation (a PMIR). It represents itself as directed towards certain aspects of the world; its phenomenal space is a perspectival space, and its experiences are subjective experiences. The intentionality relation is primarily an epistemic (but also an agentive) relation between subject and object: A mental state becomes a carrier of knowledge in virtue of being directed at something other than itself—like an arrow pointing from a person’s mind to an object in the real or even just in a possible world. What philosophers have overlooked is that intentionality is not only a theoretical problem, but that the conscious brain itself represents the intentional relation — non-conceptually, non-propositionally, subsymbolically and dynamically. In short, it is one thing to be a biological organism that represents the world, and it is another thing to consciously represent yourself as representing, in “real-time” and while this is actually happening. Now, there is not only a neurobiologically anchored self-model, but also a dynamic phenomenal simulation of the self as subject embedded in the world via constantly changing epistemic relations and agentive interactions (a system directed at variable sets of truth-conditions and satisfaction conditions). The PSM has turned into an EAM, an “epistemic agent model”. A conscious human being is a system that is capable of dynamically co-representing the representational relation while representational acts are taking place, and the instrument it uses for this purpose is the PMIR. The PMIR, the phenomenal model of the intentionality relation, is just another naturally evolved virtual organ, just like the PSM. At the end of this contribution, the proposal for epistemic (as opposed to phenomenal) subjectivity is as follows: Subjective knowledge is knowledge represented under a PMIR, that is, knowledge represented in a highly specific neurobiological data-format. What, today, we call a “first-person perspective” is a naturally evolved epistemic modality, formally characterized by this data-format.*

**Keywords:** *consciousness, self-consciousness, first-person perspective, self-model, phenomenal transparency, phantom limbs, robotics, out-of-body experience, rubber-hand illusion, evolution of tool-use*