

Markus F. Peschl (Hg.)

*Die Rolle der Seele in der Kognitionswissenschaft und der Neurowissenschaft  
Auf der Suche nach dem Substrat der Seele*

Würzburg: Königshausen & Neumann 2004

## Selbstbewusstsein in kognitiven Systemen<sup>1</sup>

Ansgar Beckermann

### 1.

Etwas provokant könnte man die Ausgangsfrage meiner Überlegungen so formulieren: „Welche kognitiven Systeme haben ein Selbst?“ Oder: „Können auch künstliche kognitive Systeme ein Selbst haben?“ Doch die Rede von *einem* oder *dem* Selbst ist höchst fragwürdig. Das Wort ‚selbst‘ ist in erster Linie ein Pronomen wie in den Sätzen „Er kam selbst“ oder „Ich weiß nicht, ob der Bundespräsident selbst kommen wird“. Es kann auch ein Adverb sein wie in „Selbst er konnte die Niederlage nicht verhindern“. Das Substantiv ‚Selbst‘ kommt dagegen nur in bestimmten Redewendungen vor wie in „Da zeigte sich sein wahres Selbst“. Die Überhöhung dieses Substantivs, die sich in Sätzen wie „Das Selbst jedes Menschen ist unzerstörbar“ zeigt, ist dagegen eine philosophische Erfindung, die wohl im Wesentlichen auf John Locke zurückgeht. Im zweiten Buch des *Essay Concerning Human Understanding* jedenfalls schreibt Locke:

Self is that conscious thinking thing, (whatever Substance made up of whether Spiritual or Material, Simple or Compounded, it matters not), which is sensible, or conscious of Pleasure and Pain, capable of Happiness or Misery, and so is concern'd for it *self*, as far as that consciousness extends. (ECHU II xxvii 17)

Der linguistische Hintergrund dieser ungewöhnlichen Verwendung von ‚Selbst‘ bzw. ‚self‘ ist offenbar die Tatsache, dass man früher im Englischen Wörter wie ‚my self‘ oder ‚it self‘ auseinander schreiben konnte, was natürlich die Vermutung zumindest begünstigt, es gäbe da so etwas wie mein Selbst. Dass diese Lesart jedoch zumindest eigenartig ist, zeigt sich sehr deutlich in einer neueren englischen Ausgabe von Descartes' *Meditationen*, wo man als Übersetzung des eigentlichen harmlosen Satzes „Nunquid *me ipsum* non tantum multo verius, multo certius, sed etiam multo distinctius evidentiusque, cognosco?“ (Descartes *Meditationes* 33 – Hervorhebung vom Verf.) Folgendes finden kann: „Surely my awareness of *my own self* is not merely much truer and more certain but also much more distinct and evident.“<sup>2</sup>

<sup>1</sup> Bei diesem Aufsatz handelt es sich um die deutsche Fassung von Beckermann 2003. Für äußerst hilfreiche Kommentare zu einer früheren Fassung möchte ich Christian Nimtz herzlich danken.

<sup>2</sup> Descartes, René *Selected Philosophical Writings* (86 – Hervorhebung vom Verf.). In der Haldane/Ross Ausgabe von 1911 hieß es noch: „[D]o I not know *myself*, not only with

Auf der anderen Seite ist die ungewöhnliche neue Verwendung des Wortes ‚Selbst‘ bzw. ‚self‘ bei Locke zwar ärgerlich, aber nicht unbedingt schädlich; denn Locke sagt ja ausdrücklich, was er unter einem Selbst verstehen will: Ein Selbst ist das Ding, das Lust und Schmerz (bewusst) erfahren kann, das zu Glück und Unglück fähig ist und das insofern um sich selbst besorgt ist. Das Selbst scheint also gar nichts anderes zu sein als der Mensch oder die Person.<sup>3</sup> Und Locke lässt bewusst offen, ob dieses Selbst etwas Materielles oder etwas (immateriell) Geistiges ist. Er will also bzgl. des Cartesischen Dualismus keine Stellung beziehen, lässt aber ausdrücklich doch die Möglichkeit offen, dass das eigentliche Selbst (!) des Menschen eine Cartesische *res cogitans* ist. Es ist daher nicht verwunderlich, dass das Reden von einem ‚Selbst‘ oft mit impliziten oder expliziten Cartesischen Konnotationen verbunden ist.

Trotzdem: Ein Selbst ist nach Locke einfach das, was zu angenehmen und unangenehmen Empfindungen und damit zu Glück und Unglück fähig ist. In der neueren Literatur steht neben diesem jedoch häufig noch ein anderer Aspekt im Vordergrund. So schreibt z.B. E.J. Lowe:

[Selves] are subjects of experience which have the capacity to recognise themselves as being individual subjects of experience. Selves possess reflexive self-knowledge. By ‚reflexive self-knowledge‘ I mean, roughly speaking, knowledge of one’s own identity and conscious mental states – knowledge of who one is and of what one is thinking and feeling. ... [R]oughly speaking – having the kind of reflexive self-knowledge which makes one a person goes hand-in-hand with possessing a ‚first-person‘ concept of oneself, the linguistic reflection of which resides in an ability to use the word ‚I‘ comprehendingly to refer to oneself. (Lowe 2000, 264f.)

Außer durch die Fähigkeit, Erfahrungen machen zu können, sind Selbste (diesen Plural akzeptiert selbst der Duden nicht!) nach Lowe also dadurch charakterisiert, dass sie reflexive Selbstkenntnis besitzen – und d.h. offenbar im besonderen Wissen, dessen Inhalt der Wissende selbst adäquat nur unter Verwendung des Wortes ‚Ich‘ formulieren kann. So verstanden sollte man sich durch das Wort ‚Selbst‘ nicht zu Fragen verführen lassen wie „Welche kognitiven Systeme haben ein Selbst?“. Denn erstens müsste die Frage eigentlich lauten: „Welche kognitiven Systeme sind Selbste?“. Und zweitens lässt sich der Inhalt dieser Frage auch in normalem Deutsch formulieren: „Welche kognitiven Systeme verfügen über reflexives Wissen, dessen Inhalt adäquat nur unter Verwendung des

---

much more truth and certainty, but also with much more distinctness and clearness?“ (156 – Hervorhebung vom Verf.).

<sup>3</sup> Vgl. z.B. die folgende entlarvende Bemerkung Disraelis: „Self is the only *person* whom we know anything about.“ *The Oxford English Dictionary*, 2nd ed., s.v. ‚self‘, C.I.1.e. (Hervorhebung vom Verf.).

Wortes ‚Ich‘ formuliert werden kann?“<sup>4</sup> Dieser durchaus vernünftigen Frage, die nicht gleich insinuiert, dass Selbste eine besondere, mystische Art von Entitäten darstellen, will ich im Folgenden nachgehen.

## 2.

Kognitive Systeme sind Systeme, die versuchen, sich ein Bild von der Welt zu machen, in der sie leben. Sie repräsentieren ihre Umwelt, um in dieser Umwelt besser zurechtzukommen. Grundsätzlich können diese Repräsentationen sehr unterschiedliche Formen annehmen. Im Folgenden werde ich allerdings nur solche kognitive Systeme betrachten, die ihre Umwelt explizit in Form von Listen, also in einer bestimmten Spielart einer *lingua mentis*, repräsentieren. Meine These ist, dass solche kognitiven Systeme genau dann über die für ein Selbst erforderliche reflexive Selbstkenntnis verfügen, wenn sie über eine ganze bestimmte Art von Repräsentationen verfügen – Repräsentationen, die von ihnen selbst handeln und die darüber hinaus auf eine ganz bestimmte Weise von ihnen selbst handeln. Über kognitive Systeme anderer Art werde ich hier nichts sagen. Allerdings bin ich davon überzeugt, dass auch diese anderen kognitiven Systeme genau dann über die für ein Selbst erforderliche reflexive Selbstkenntnis verfügen, wenn es ihnen *analoge* Strukturen oder Fähigkeiten gibt. Ich bin also davon überzeugt, dass sich die folgenden Überlegungen – mit mehr oder weniger großem Aufwand – auf andere kognitive Systeme übertragen lassen. Mehr will ich darüber hier aber nicht sagen. Konzentrieren wir uns also auf die Frage: Welche Repräsentationen müssen die von mir behandelten kognitiven Systeme aufbauen, um über die für ein Selbst erforderliche reflexive Selbstkenntnis zu verfügen?<sup>5</sup>

John Perry hat in „Myself and I“ eine für diese Frage äußerst wichtige Unterscheidung zwischen drei Arten selbstbezogenen Wissens getroffen – die Unterscheidung zwischen akteurzentriertem Wissen (*agent-relative knowledge*), eigentlichem Selbstwissen (*self-attached knowledge*) und Wissen, das sich auf eine Person bezieht, die (zufällig) man selbst ist (*knowledge of the person one happens to be*). Akteurzentriertes Wissen liegt dann vor, wenn ein System die Umwelt von seiner eigenen Perspektive aus repräsentiert. Es setzt nicht voraus, dass dieses System, der Akteur, über Repräsentationen verfügt, die sich explizit auf es selbst beziehen. Das System braucht keinen Begriff von sich selbst zu haben. In unserem Kontext bedeutet das: Die Listen, in denen akteurzentriertes Wissen repräsentiert wird, müssen keine Symbole enthalten, die sich explizit auf das System beziehen. Letzten Endes besteht akteurzentriertes Wissen also in Repräsen-

---

<sup>4</sup> Genauer muss es natürlich heißen: „Welche kognitiven Systeme verfügen über reflexives Wissen, dessen Inhalt adäquat nur unter Verwendung von Quasi-Indikatoren im Sinne Castañedas formuliert werden kann?“

<sup>5</sup> Die Geschichte, mit der ich diese Frage beantworten werde, ist natürlich nicht neu. Vgl. z.B. Rosenberg 1986, bes. Kap. VI und VII.

tationen, die in der Kognitionswissenschaft schon lange bekannt sind – Repräsentationen, in denen die Umwelt nicht in Welt-, sondern in beobachterzentrierten Koordinaten repräsentiert wird. Was das heißt, lässt sich an folgendem Beispiel gut veranschaulichen.

Eines der frühen Erfolgsprogramme der KI war das Programm SHRDLU – ein Programm, in dem Terry Winograd versuchte, Sprachverstehen und Handlungsplanung in ein System zu integrieren. SHRDLU ‚lebt‘ in einer Art Mikrowelt, in der es eine Reihe unterschiedlicher Gegenstände gibt – Blöcke, Kugeln, Pyramiden und eine Schachtel. Die Aufgaben des Programms bestehen in der Regel darin, diese Gegenstände auf bestimmte Weise neu anzuordnen – also z.B. die grüne Pyramide auf den roten Block zu stellen oder die Kugel in die Schachtel zu legen. Für uns ist hier nur wichtig, wie SHRDLU seine ‚Umwelt‘ repräsentiert. Die Repräsentation einer Situation wie der, die in Abbildung 1 dargestellt ist, hat z.B. die folgende Form:

(Ist-ein	Objekt-1	Block)
(Farbe	Objekt-1	grün)
(Ort	Objekt-1	(1 1 2))
(Größe	Objekt-1	(2 2 2))
...		
(Ist-ein	Objekt-5	Kugel)
(Farbe	Objekt-5	grün)
(Ort	Objekt-5	(4 3 0))
(Größe	Objekt-5	(2 2 2))
...		
(Hält	Hand	nichts)
(Ort	Hand	(2 5 7))

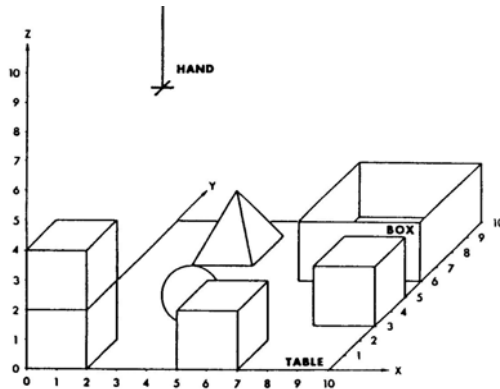


Abbildung 1

Entscheidend ist hier, dass alle Informationen über den Ort der beteiligten Objekte in so genannten *Weltkoordinaten* repräsentiert sind – sogar der Ort, an dem sich die Greifhand des Systems befindet. Wenn SHRDLU einen Gegenstand *a* greifen will, muss es also überlegen: Wo befindet sich *a*? Wo befindet sich die Hand? Wie kommt die Hand von dort zum Ort von *a*? Das erscheint nicht nur extrem unnatürlich. Es stellt sich auch die Frage, wie kognitive Systeme eigentlich Wissen darüber erwerben sollen, an welchen *objektiven* Koordinaten sich die Gegenstände in ihrer Umgebung und sie selbst befinden. Wir Menschen jedenfalls nehmen die Welt ganz offensichtlich nicht in Weltkoordinaten wahr. Wenn ich von der Tür aus in mein Arbeitszimmer schaue, sehe ich vielmehr ungefähr folgendes:

Ca. 5 Schritte geradeaus befindet sich ein Schreibtisch.  
 Direkt hinter dem Schreibtisch ist ein Fenster.  
 Links auf dem Schreibtisch steht ein Bildschirm.  
 Davor eine Tastatur.  
 Mitten auf dem Schreibtisch steht eine Tasse.  
 Links neben dem Schreibtisch ist ein Regal.  
 In dem Regal steht halbhoch ein Drucker.

Ich repräsentiere den Ort der Dinge in meiner Umgebung also durch die räumlichen Relationen, in denen sie *zu mir* und *zueinander* stehen. Und dafür gibt es zwei sehr gute Gründe. Zum ersten dieser Gründe schreibt Perry:

Everything we learn about other objects we learn by employing methods that are appropriate because those objects stand in certain relations to us. ... [The objects in our vicinities have] *agent-relative roles*: roles that other individuals play in the lives of agents. These are agent-relative roles, because an object plays or doesn't play such a role relative to a given agent, at a given time. For example, my computer is playing the role of *object in front* right now, relative to me, but not relative to you. ... This is the first of two very general facts I want to emphasize: any object we learn about plays some agent-relative role, basic or derived, in our life. We learn about the object by using an epistemic method connected to the role, a way of finding out about the object or person playing that role. The way to find out about the object in front of you is to look at it, or perhaps to walk up to it and touch it. (Perry 1998, 84f.)

Mit anderen Worten: Wir können nur etwas über die Dinge unserer Umwelt erfahren, weil sie zu uns in bestimmten Beziehungen stehen, weil sie uns gegenüber bestimmten Rollen einnehmen. Es ist daher kein Wunder, dass wir diese Dinge zunächst als Träger dieser Rollen repräsentieren – als den Tisch 5 Schritte vor uns, die Person rechts neben uns, der Boden unter unseren Füßen, usw. Noch wichtiger ist aber ein anderer Punkt. Nur wenn ich die Dinge meiner Umwelt als Träger der Rollen repräsentiere, die sie mir gegenüber spielen, weiß ich unmittelbar, was von dem, was um mich herum vorgeht, für mich relevant ist. Nehmen wir einen Ball, der auf mich zufliegt. Wenn ich die Bewegung des

Balles als ein Auf-mich-Zufliegen repräsentiere, folgt unmittelbar, dass ich etwas tun muss – den Ball auffangen oder mich bücken oder was immer. Wenn ich die Bewegung des Balles dagegen in Weltkoordinaten repräsentiere, folgt daraus nur dann etwas, wenn ich weiß, an welchen Weltkoordinaten ich mich selbst befinde und ob dieser Ort auf der Bahn des Balles liegt. Offenbar erleichtern akteurzentrierte Repräsentationen es uns also erheblich, das, was um uns herum passiert, in seiner Handlungsrelevanz richtig einzuschätzen. Dies führt auch sofort zu dem zweiten Grund, über den Perry schreibt:

... [E]verything we do comes down to performing operations on the objects around us – objects in front of us, behind us, above us; objects we are holding; objects we can see. By doing these things, we do things to objects in less basic relations to us. ... I know how to move my body so as to effect objects around me, and I know how effecting those objects will effect other objects related to them in certain ways. (Perry 1998, 85)

Wenn wir handeln, führen wir immer Körperbewegungen mit Bezug auf Dinge aus, die uns gegenüber eine bestimmte Rolle einnehmen. Es gibt einen bestimmten Typ von Körperbewegung, mit dem wir eine Tasse greifen, die vor uns auf dem Tisch steht; es gibt einen bestimmten Typ von Körperbewegung, mit dem wir eine Fliege von unserer Wange verscheuchen; und es gibt – Perrys Lieblingsbeispiel – einen bestimmten Typ von Körperbewegung, mit dem wir die Person links neben uns erstechen. Aber es gibt *keinen* bestimmten Typ von Körperbewegung, mit dem wir die Pyramide am Ort (a b c) auf den Block am Ort (d e f) setzen. D.h., wenn wir überlegen, mit welchem Typ von Körperbewegung wir eine bestimmte Handlung vollziehen können, ist es überaus hilfreich, wenn die Orte der Dinge in unserer Umwelt in akteurzentrierten Koordinaten repräsentiert sind. Ich weiß, mit welcher Körperbewegung ich eine Tasse von mir auf dem Schreibtisch greifen kann; aber ich weiß nicht ohne weiteres, mit welcher Körperbewegung ich eine Tasse greifen kann, die sich an der Stelle (7 10 3) befindet.

There are then two kinds of methods connected with agent-relative roles, epistemic methods and pragmatic methods. These two kinds of methods are the key to all human intelligence and purposive activity. We know how to find out what kinds of objects occupy these roles, and we know how to perform various operations on them. ... Our practical knowledge then, the knowledge that enables us to do things, forms a structure at whose base is information about the objects that play relatively basic agent-relative roles in our lives. (Perry 1998, 85)

Warum ist das für uns wichtig? Nun, weil akteurzentrierte Repräsentationen immer einen Bezug auf den Akteur selbst beinhalten. Nicht ohne Grund sprechen Kognitionswissenschaftler in diesem Zusammenhang von einem ‚ego-centric reference system‘. Wenn ich einen Apfel auf dem Tisch liegen sehe, in die Hand nehme und aufesse, beruhen die komplexen Körperbewegungen, mit deren Hilfe ich den Apfel in meinen Mund befördere, darauf, dass ich durch Wahr-

nehmung gelernt habe, dass sich der Apfel in einer bestimmten räumlichen Beziehung zu *mir* befindet. Wenn ich die Tastatur meines Computers nutze, muss ich meine Finger in einer bestimmten Entfernung und Richtung von *mir* bewegen. „It isn’t enough to know where the buttons were relative to one another, or where the [keyboard] was in the building or the room. I had to know where these things were relative to *me*. It seems then, that these basic methods already require me to have some notion of myself.“ (Perry 1998, 86)

Schon akteurzentrierte Repräsentationen scheinen also zu den Repräsentationen zu zählen, deren Inhalt nur unter Verwendung des Wortes ‚ich‘ sprachlich angemessen wiedergegeben werden kann. Doch dieser Schein trügt. Wie auch Perry betont, ist das Wort ‚ich‘ bei der sprachlichen Formulierung des Inhalts dieser Repräsentationen immer verzichtbar. Wenn ich die Situation repräsentieren möchte, dass sich 50 cm vor mir leicht nach rechts versetzt ein Apfel befindet, dann muss ich nicht repräsentieren, dass sich der Apfel in einer bestimmten räumlichen *Relation* zu *mir* befindet, in der er sich auch zu anderen Objekten befinden könnte. Vielmehr geht es mir in diesem Fall nur darum, dass der Apfel eine bestimmte Eigenschaft hat – nämlich die Eigenschaft, sich in einer bestimmten räumlichen *Relation-zu-mir* zu befinden. Da das zweite Relationsglied – ich – in all diesen Fällen gleich bleibt, kann ich es sozusagen festhalten und mich auf die Frage konzentrieren, welche Dinge in meiner Umwelt haben die Eigenschaft, in dieser Relation-zu-mir zu stehen. Es ist so, als würde ich die Situation, dass sich 50 cm vor mir leicht nach rechts versetzt ein Apfel befindet, in Polarkoordinaten  $r$  und  $\alpha$  repräsentieren:

(Ist-ein Objekt-A Apfel)  
(Ort Objekt-A (0.5 8°))

Auch in einer solchen Repräsentation wird ganz offensichtlich nicht auf den Ursprung des Koordinatensystems – auf mich – Bezug genommen. Diese Bezugnahme ist, wie gesagt, verzichtbar. Perry selbst drückt das so aus:

The general point is this. Sometimes all of the facts we deal with involving a certain  $n$ -ary relation involve the same object occupying one of the argument roles. In that case, we don’t need to worry about that argument role; we don’t need to keep track of its occupant, because it never changes. (Perry 1998, 87)

Akteurzentrierte Repräsentationen stellen also eine Form von Selbstkenntnis dar; sie beinhalten Wissen darüber, wie sich die Dinge um den Akteur herum zu ihm selbst verhalten. Insofern beinhalten sie immer auch Wissen über den Akteur selbst. Allerdings setzen diese Repräsentationen doch noch nicht voraus, dass der Akteur einen *Begriff von sich* hat. Denn sie müssen keinen Ausdruck enthalten, der sich explizit auf den Akteur bezieht. Dies gilt im Übrigen auch für alle Repräsentationen, die sich auf Eigenzustände des Systems beziehen. Empfindungen etwa können auf die folgende Weise repräsentiert werden:

Schmerzen im linken Knie.  
Leichtes Kribbeln in der Magengegend.  
Jucken auf dem Kopf.

Auch hier ist der explizite Bezug auf das System verzichtbar. Denn in dieser Form wird ein System Empfindungen niemals anderen Wesen, sondern immer nur sich selbst zuschreiben. Also muss es selbst nicht eigens erwähnt werden.

Erst Repräsentationen, die Selbstkenntnis im strengen Sinne (*self-attached knowledge* im Sinne Perrys) darstellen, erfordern, dass das System einen *Begriff von sich* entwickelt. Was sind das für Repräsentationen und wie können sie entstehen?

### 3.

Kommen wir noch einmal auf den Grundsatz zurück: Kognitive Systeme sind Systeme, die versuchen, sich ein Bild von der Welt zu machen, in der sie leben, d.h. Systeme, die versuchen, ihre Umwelt zu repräsentieren. Wie geht das vor sich? Was ist dafür erforderlich, dass es einem kognitiven System – nennen wir es ‚AL‘ – gelingt, seine Umwelt zu repräsentieren? Welche Prozesse sind hier beteiligt?

Offenbar ist es zunächst so, dass die Umwelt in AL kausale Spuren hinterlässt und dass AL das Problem lösen muss, aus diesen Spuren seine Umwelt zu rekonstruieren. Genauer heißt das, dass es AL durch eine Auswertung der kausalen Spuren gelingen muss, die folgenden Fragen zu beantworten:

1. Wie viele *Objekte* gibt es in der aktuellen Szene?
2. Zu welcher *Art* von Dingen gehören diese Objekte?
3. *Wo* befinden sich diese Objekte?
4. Welche *Eigenschaften* haben sie und in welchen *Relationen* stehen sie zueinander?

Wenn AL diese Fragen beantwortet hat, kann er auf die folgende Weise Repräsentationen seiner Umwelt aufbauen.

- Nach der Beantwortung der ersten Frage gibt AL jedem der an der Szene beteiligten Objekte einen internen Namen (etwa ‚Objekt-36‘, ‚Objekt-37‘, usw.).
- Nach Beantwortung der zweiten Frage fügt AL der Liste seiner Repräsentationen für jedes Objekt eine Repräsentation der Form  
(Ist-ein Objekt-*x* *Typ*)  
hinzu.
- Nach Beantwortung der dritten Frage fügt AL der Liste seiner Repräsentationen für jedes Objekt eine Repräsentation der Form  
(Ort Objekt-*x* *Koordinaten*)  
hinzu.

- Nach der Beantwortung der vierten Frage schließlich fügt AL der Liste seiner Repräsentationen für jedes Objekt geeignete Repräsentationen der Form  
(Farbe Objekt-*x* *Farbe*)  
(Größe Objekt-*x* *Größe*)  
(*Relation* Objekt-*x* Objekt-*y*)  
usw.

hinzu.

Soweit die Grundidee. Aber das alleine reicht häufig noch nicht aus. Viele kognitive Systeme müssen darüber hinaus das Problem lösen, Objekte *wiederzuerkennen*, denen sie früher schon einmal begegnet sind. Wenn sich ein anderer Storch seinem Nest nähert, muss der Storch im Nest nicht nur entscheiden: Ist das ein weiblicher oder ein männlicher Storch? Er muss auch eine Antwort auf die Frage finden: Ist das meine Partnerin oder ein fremder Storch? Wenn AL bei der Szenenanalyse jedem Objekt, das er wahrnimmt, einen *neuen* internen Namen gäbe, wäre das aber so, als wäre ihm dieses Objekt noch nie begegnet, als wäre dieses Objekt völlig neu für ihn. Doch das würde im Zweifelsfall zu ganz und gar unangemessenen Reaktionen führen. Also muss AL auch die Aufgabe lösen, – etwa durch Vergleich von typischen Merkmalen – herauszufinden, welche der in der analysierten Szene vorgefundenen Objekte mit Dingen identisch sind, denen schon einmal begegnet ist und denen er daher schon früher einen internen Namen gegeben hat. Für diese Objekte muss AL in allen neuen Repräsentationen den schon vorhandenen Namen verwenden.<sup>6</sup>

Für uns ist aber folgendes von besonderer Bedeutung: Bis jetzt gab es für AL noch keinerlei Grund, in den gespeicherten Repräsentationen einen Namen für sich selbst zu verwenden. Es kann durchaus sein, dass AL den Ort, an dem sich die von ihm wahrgenommenen Dinge befinden, in akteurzentrierten Koordinaten repräsentiert und dass AL repräsentiert, in welchen Relationen-zu-ihm sich diese Dinge befinden. Doch all dies erfordert, wie wir schon gesehen haben, keineswegs, dass die entsprechenden Repräsentationen einen Ausdruck enthalten, der sich explizit auf AL bezieht. Daher ist die Frage jetzt: Was muss passieren, damit es für AL sozusagen unvermeidlich wird, einen solchen Ausdruck zu verwenden? Diese Frage lässt sich, wie mir scheint, am besten beantworten, wenn wir ALs kognitive Geschichte noch ein wenig weiter erzählen.

Der nächste wichtige Schritt beruht auf der Beobachtung, dass in der Umgebung kognitiver Wesen häufig andere kognitive Wesen vorkommen. Und auch von AL wollen wir annehmen, dass es in seiner Umgebung andere kognitive Systeme gibt, denen er von Zeit zu Zeit begegnet. Das Problem mit diesen kog-

<sup>6</sup> Perry geht in entsprechenden Überlegungen nicht davon aus, dass Objekte interne Namen erhalten, sondern dass für jeden Gegenstand ein Ordner angelegt wird, in dem alle Informationen über diesen Gegenstand gesammelt werden. Für ihn stellt sich das Problem daher so dar: Muss AL für einen Gegenstand, den er bei der Analyse einer Szene entdeckt, einen neuen Ordner anlegen oder kann er die neu gewonnenen Informationen über diesen Gegenstand in einem schon vorhandenen Ordner ablegen. (Cf. Perry 1998, 89ff.)

nitiven Wesen ist, dass ihr Verhalten nicht nur von ihren ‚natürlichen‘ Eigenschaften abhängt, sondern ganz entscheidend auch davon, welches Bild *sie* sich von der Umwelt machen, wie *sie* die Welt repräsentieren. Wenn AL das Verhalten seiner Mitwesen richtig voraussagen möchte, bleibt ihm daher gar nicht anderes übrig als Repräsentationen zu entwickeln, die die Repräsentationen seiner Kumpane zum Gegenstand haben – Repräsentationen, die man gemeinhin ‚Metarepräsentationen‘ nennt. Diese Repräsentationen haben generell die Form:

(Glaubt Objekt-x Repräsentation)

oder

(Wünscht Objekt-x Repräsentation)

usw.

Es ist eine interessante Frage, wie die Repräsentationen, die für die Variable ‚Repräsentation‘ eingesetzt werden können, im Einzelnen aussehen. Manchmal wird AL zu der Überzeugung kommen, dass eines seiner Mitwesen – etwa das Wesen mit dem internen Namen ‚Objekt-111‘ – eine Überzeugung über einen Gegenstand hat, der AL selbst bekannt ist und für den er intern den Namen ‚Objekt-7‘ verwendet. Dann würde die entsprechende Metarepräsentation so aussehen:

(Glaubt Objekt-111 (Farbe Objekt-7 grün)).

Mit dieser Metarepräsentation würde AL seinem Mitwesen sozusagen eine *de re* Überzeugung zuschreiben; denn diese Repräsentation hätte den Inhalt: Objekt-111 glaubt *von dem Gegenstand*, den AL unter dem Namen ‚Objekt-7‘ kennt, dass er grün ist.

Es kann aber auch vorkommen, dass ALs Mitwesen auf etwas starrt, das AL nicht sehen kann, und dass das Mitwesen sich dabei so verhält, wie es sich sonst nur verhält, wenn es eine Spinne sieht. Wie wird AL diese Situation repräsentieren? Nun, erstens sieht AL das Ding, das sein Mitwesen offenbar vor sich hat, nicht selbst. Also muss AL einen *neuen* internen Namen verwenden – sagen wir ‚Objekt-57‘. Zweitens weiß AL von dem durch diesen neuen Namen bezeichneten Gegenstand nichts – außer, dass er sich direkt vor seinem Mitwesen befindet und dass sein Mitwesen offenbar glaubt, dass dieser Gegenstand eine Spinne ist. Also wird AL seinem Repräsentationssystem die folgenden beiden Listen hinzufügen:

(Vor Objekt-57 Objekt-111)

(Glaubt Objekt-111 (Ist-ein Objekt-57 Spinne)).

Es kann natürlich auch noch sein, dass sich AL die Überzeugung seines Mitwesens zu eigen macht und daher zusätzlich die Repräsentation ausbildet

(Ist-ein Objekt-57 Spinne).

Aber das steht hier nicht zur Debatte. Wichtig ist nur, dass AL, wenn er Überzeugungen seiner Kumpane repräsentieren will, die sich auf Dinge beziehen, die er nicht kennt, *neue* interne Namen verwenden muss.

Und wichtig ist darüber hinaus, dass AL gut daran tut, auch die Empfindungen seiner Mitwesen zu repräsentieren – etwa so:

(Schmerzen-im-Knie Objekt-111).

Denn auch Schmerzen, Freude und andere Empfindungen sind ja verhaltensrelevant. Allerdings: Der wichtigste Schritt kommt erst noch.

Irgendwann nämlich wird AL darauf kommen, dass er für die anderen kognitiven Wesen nichts anderes ist als sie für ihn. Diese anderen Wesen repräsentieren ihre Umwelt nämlich so, dass in ihr ein Wesen vorkommt, das seinerseits die Welt repräsentiert und das *de facto* niemand anderes ist als – AL. Irgendwann wird AL etwa merken, dass eines seiner Mitwesen IHN anstarrt oder sich IHM nähert oder etwas von IHM will, z.B. Essen. Und: Irgendwann wird es nicht mehr ausreichen, solche Situationen auf akteurzentrierte Weise zu repräsentieren. Mit anderen Worten: Um repräsentieren zu können, welche Repräsentationen seine Mitwesen über ihn, AL, haben, muss AL einen internen Namen – sagen wir ‚Objekt-100‘ – für sich selbst verwenden. Erst mit Hilfe dieses Namens kann er nämlich den Wunsch seines Mitwesens angemessen repräsentieren:

(Wünscht Objekt-111 (Gibt-Essen Objekt-100 Objekt-111)).

Ein ähnlicher Fall würde eintreten, wenn AL merkt, dass das Wesen mit dem Namen ‚Objekt-111‘ offenbar von ihm selbst, AL, glaubt, er habe Schmerzen im Knie; dies müsste AL nämlich so repräsentieren:

(Glaubt Objekt-111 (Schmerzen-im-Knie Objekt-100)).

Offenbar gibt es keine Möglichkeit, die Überzeugung seines Mitwesens auf akteurzentrierte Weise zu repräsentieren. AL benötigt also einen internen Namen für sich selbst – einen Namen, der zunächst allerdings nur ein Name ist wie jeder andere. Irgendwann wird es AL aber wohl dämmern, dass der Ausdruck ‚Objekt-100‘ ein Name für ihn selbst, AL, ist. Sicher, das ist nur eine metaphorische Ausdruckweise. Denn wie kann AL merken, dass ‚Objekt-100‘ ein Name für ihn selbst ist, wenn er noch gar keinen Begriff von sich selbst hat? Wir müssen also herausfinden, was mit dieser Metapher gemeint sein kann, was es heißen kann, dass AL merkt, dass ‚Objekt-100‘ ein Name für ihn selbst ist – allgemeiner: was es heißen kann, dass AL einen Begriff von sich ausbildet. Drei Schritte sind hier von besonderer Bedeutung:

1. AL beginnt, in seinen Repräsentationen einen neuen Namen zu verwenden, wenn es darum geht, Repräsentationen anderer kognitiver Wesen zu repräsentieren, die AL selbst zum Gegenstand haben.
2. AL beginnt damit, diesen Namen auch zu verwenden, wenn er sich z.B. im Spiegel sieht.
3. In ALs kognitiver Architektur entwickelt sich eine systematische Verbindung zwischen Repräsentationen, die diesen neuen Namen enthalten, und den schon früher erzeugten akteurzentrierten Repräsentationen, in denen nur implizit auf AL Bezug genommen wird.

Dieser letzte Punkt ist entscheidend. Denn es ist wirklich von allergrößter Bedeutung, wenn in ALs kognitiver Architektur die Repräsentation

(Sitze-auf Objekt-3)

anfängt, dieselbe Rolle zu spielen wie die Repräsentation

(Sitzt-auf Objekt-100 Objekt-3),

und wenn die Repräsentation  
(Schmerzen-im-Knie)  
beginnt, dieselbe Rolle zu spielen wie die Repräsentation  
(Schmerzen-im-Knie Objekt-100).

Das Ergebnis dieses Prozesses ist, dass sich in ALs kognitiver Architektur eine Äquivalenz entwickelt zwischen älteren akteurzentrierten Repräsentationen und Repräsentationen, die sich mit Hilfe des neuen Namens ‚Objekt-100‘ explizit auf AL beziehen. Und das bedeutet auch, dass von diesem Zeitpunkt an der gesamte Input aus ALs Körper – d.h., alles was ihm über Propriozeption zugänglich ist – nicht nur zu akteurzentrierten Repräsentationen, sondern auch zu Repräsentationen führt, die sich explizit auf AL beziehen. Mit anderen Worten: AL entwickelt ein Körperschema.

Früher repräsentierte AL seine Umgebung anhand der Frage: Ich welcher Relation-zu-mir steht das Objekt  $x$ ? Und der Grund dafür war, dass das zweite Relationsglied – AL selbst – in allen Fällen gleich blieb. Jetzt macht er diese Fixierung rückgängig, weil sich herausgestellt hat, dass er nur eines von vielen Dingen ist, die in derselben Relation zu  $x$  stehen können. Metaphorisch: AL beginnt sich mit den Augen der anderen zu sehen.

Dies hat noch eine weitere Wirkung: AL beginnt, Metarepräsentationen mit Bezug auf sich selbst zu entwickeln. Früher war es für ihn einfach nicht nötig zu wissen, was er selbst glaubt und wünscht. Jetzt jedoch, wo er beginnt, sich mit den Augen der anderen zu sehen, wird alles anders. Denn das Verhalten der anderen hängt auch davon, was sie darüber denken, wie er, AL, die Welt repräsentiert. Also muss er anfangen, sich um seine eigenen intentionalen Zustände zu kümmern. Und das führt schließlich dazu, dass AL beginnt, die für ein Selbst notwendige Selbstkenntnis zu entwickeln. Denn die hatte Lowe ja so charakterisiert:

[By] ‚reflexive self-knowledge‘ I mean, roughly speaking, knowledge of one’s own identity and conscious mental states – knowledge of who one is and of what one is thinking and feeling. (Lowe 2000, 264f.)

#### 4.

Aber, so könnte man einwenden, der Name ‚Objekt-100‘ ist doch ein interner Name wie jeder andere. Wie kann es sein, dass Repräsentationen, in denen dieser Name vorkommt, einen so speziellen Status haben, dass sich gerade aus diesen Repräsentation die Art von Selbstkenntnis ergibt, die dafür sorgt, dass AL über Selbstbewusstsein verfügt oder – in Lowes Redeweise – dass AL ein Selbst ist?

Nun, zunächst ist sicher vorstellbar, dass AL zusammen mit seinen Kumpanen beginnt, eine Sprache zu entwickeln, in der auch indexikalische Ausdrücke wie ‚ich‘, ‚du‘, ‚dort‘ usw. vorkommen. (Vielleicht ist das sogar notwendig dafür, dass AL die entsprechenden Repräsentationen und Metarepräsentationen entwi-

ckelt.) Weiter ist auch vorstellbar, dass AL lernt, nur solche Repräsentationen, in denen der Name ‚Objekt-100‘ vorkommt, mit ‚ich‘-Sätzen auszudrücken. Aber wäre das nicht ein bloßer Zufall? Ist nicht ebenso gut vorstellbar, dass AL sich angewöhnt, nur Repräsentationen, in denen der Name ‚Objekt-13‘ vorkommt, mit ‚Ich‘-Sätzen auszudrücken, wobei ‚Objekt-13‘ ein Name für einen beliebigen anderen Gegenstand ist? Die Antwort lautet: Nein. Den Gebrauch des Wortes ‚ich‘ zu lernen heißt unter anderem zu lernen, dass sich jedes Mitglied der Sprachgemeinschaft mit diesem Wort nur auf sich selbst beziehen kann. AL hat die Bedeutung von ‚ich‘ also nur dann gelernt, wenn er gelernt hat, mit Hilfe dieses Wortes nur Repräsentationen auszudrücken, die sich auf ihn selbst beziehen. Doch diese Antwort reicht noch nicht aus. Denn an dieser Stelle spielt Perrys Unterscheidung zwischen *self-attached knowledge* und *knowledge of the person one happens to be* eine entscheidende Rolle. Erinnern wir uns kurz an eines der einschlägigen Beispiele.

Gleich am Beginn seines Aufsatzes „The Problem of the Essential Indexical“ schreibt Perry:

I once followed a trail of sugar on a supermarket floor, pushing my cart down the aisle on one side of a tall counter and back the aisle on the other, seeking the shopper with the torn sack to tell him he was making a mess. With each trip around the counter, the trail became thicker. ... Finally it dawned on me. I was the shopper I wanted to catch. (Perry 1979, 33)

Die Pointe dieser Geschichte ist klar. Zunächst glaubte Perry, dass jemand anderes – eine Person, über die er sonst nichts wusste – mit einem Einkaufswagen durch den Supermarkt fuhr, in dem sich eine beschädigte Packung Zucker befand. Doch nach einer gewissen Zeit bemerkte er, dass diese Person niemand anderes war als er selbst. Und dies brachte ihn zu der Überzeugung, dass sich die beschädigte Packung Zucker in seinem eigenen Einkaufswagen befinden musste. Wenn man diese beiden Überzeugungen im Sinne Russells analysiert, scheinen sie denselben Inhalt zu besitzen – die singuläre Proposition ⟨hat eine beschädigte Packung Zucker im Einkaufswagen, Perry⟩. Doch die beiden Überzeugungen müssen verschieden sein; denn ihre Verhaltenskonsequenzen unterscheiden sich ganz erheblich. Die erste Überzeugung veranlasste Perry, die Person mit der beschädigten Packung Zucker im Einkaufswagen zu suchen, um sie darauf hinzuweisen, welche Unordnung sie verursache. Die zweite Überzeugung führte dagegen zu einem ganz anderen Verhalten. „I stopped following the trail around the counter and rearranged the torn sack in my cart.“ (Perry 1972, 33) Da beide Überzeugungen denselben Russellschen Inhalt haben, können sie sich aber nur in ihren Freigeschen Inhalten – in verschiedenen Arten des Gegebenseins – unterscheiden. Und: Offenbar haben wir es bei der zweiten Überzeugung mit einer ganz besonderen Art zu tun, in Perry sich selbst ‚gegeben‘ ist – einer Art, die man als ‚EGO-Art des Gegebenseins‘ bezeichnen könnte. Die Frage, die wir uns stellen müssen, lautet daher: Was hat der interne Name ‚Objekt-100‘ mit dieser EGO-Art des Gegebenseins zu tun?

Zunächst gilt es festzuhalten, dass in ALs Repräsentationen verschiedenen Arten des Gegebenseins normalerweise durch verschiedene interne Namen Rechnung getragen wird. Angenommen, AL sieht in einem Spiegel, dass hinter einem kognitiven Wesen ein Bär auftaucht, wobei er jedoch nicht erkennt, dass es sich bei diesem kognitiven Wesen *de facto* um ihn selbst handelt. Bei der Repräsentation dieser Situation muss AL für das beobachtete kognitive Wesen daher einen neuen internen Namen, sagen wir den Namen ‚Objekt-213‘, einführen. (Dasselbe gilt natürlich für den Bären, es sei denn, AL wäre genau diesem Bären früher schon einmal begegnet.) Auf diese Weise wird AL z.B. Repräsentationen wie diese entwickeln:

(Ist-ein Objekt-511 Bär)  
 (Ist-ein Objekt-213 kognitives System)  
 (Hinter Objekt-511 Objekt-213)

Den Namen ‚Objekt-100‘ würde AL nur verwenden, wenn er bemerken würde, dass sich der Bär tatsächlich hinter seinem *eigenen* Rücken befindet.

Zweitens, und das ist noch wichtiger: Verschiedenen Arten des Gegebenseins entsprechen verschiedene Arten der kognitiven Verarbeitung, d.h. verschiedene funktional/computationale Rollen. Wenn zwei interne Ausdrücke für AL kognitiv äquivalent sind, entsprechen sie also derselben Art des Gegebenseins. Und umgekehrt: Zwei Ausdrücke mit unterschiedlichen computationalen Rollen entsprechen verschiedenen Arten des Gegebenseins. Die Frage ist also: Welche spezifischen Merkmale der computationalen Rolle des Namens ‚Objekt-100‘ sind dafür verantwortlich, dass dieser Name einer EGO-Art des Gegebenseins entspricht?

Nun, die Art und Weise, in der der Name ‚Objekt-100‘ in ALs kognitiver Architektur verarbeitet wird, unterscheidet sich in der Tat ganz wesentlich von der Verarbeitung aller anderen Namen. Schließlich haben wir angenommen, dass Repräsentationen, in denen der Name ‚Objekt-100‘ vorkommt, die einzigen Repräsentationen sind, die zu akteurzentrierten Repräsentationen äquivalent sind – wobei zwei Repräsentationen genau dann ‚äquivalent‘ heißen sollen, wenn sie dieselbe funktionale/computationale Rolle besitzen. Das hat zwei sehr wichtige Konsequenzen. Erstens: Der gesamte propriozeptive Input führt in ALs kognitiver Architektur nur zu akteurzentrierten Repräsentationen und – wegen der angenommenen Äquivalenz – zu Repräsentationen mit dem Namen ‚Objekt-100‘. Selbst wenn ‚Objekt-213‘ ein Name ist, der sich – ohne dass AL das weiß – tatsächlich ebenfalls auf AL bezieht, würde dieser Name in ALs kognitiver Architektur niemals dazu verwendet, Informationen über ALs Kopfschmerzen oder die Position seiner Glieder zu speichern. Jedenfalls dann nicht, wenn diese Informationen nicht von den äußeren Sinnen, sondern aus dem propriozeptiven System ALs stammen. Die besondere Art und Weise, in der AL viele seiner eigenen Körperzustände gegeben sind, ist also nur mit dem Namen ‚Objekt-100‘ und mit keinem anderen Namen verknüpft.

Zweitens, und mindestens ebenso wichtig: Akteurzentrierte Repräsentationen haben eine charakteristische unmittelbare Handlungswirksamkeit. Denken wir noch einmal an die Repräsentation mit dem Inhalt: Da fliegt ein Ball direkt auf mich zu. Diese Repräsentation wird mich sofort zum Handeln veranlassen. Ich werde mich ducken oder versuchen, den Ball zu fangen, oder was auch immer. Ganz anders bei einer Repräsentation mit dem Inhalt: Da bewegt sich ein Ball mit der Geschwindigkeit  $v$  von Ort  $a$  zu Ort  $b$ . Diese Repräsentation bewirkt alleine wahrscheinlich gar nichts. Erst wenn außerdem klar ist, dass ich selbst mich zwischen  $a$  und  $b$  befinde, wird mich das wahrscheinlich zum Handeln veranlassen. Und: Wie wir gesehen haben, besitzen akteurzentrierte Repräsentationen darüber hinaus ein weiteres charakteristisches Merkmal. Wenn ein kognitives System die Gegenstände in seiner Umwelt mit Hilfe der Relationen repräsentiert, in denen diese Gegenstände zu ihm *selbst* stehen, gibt es für eine Vielzahl von Handlungen klar definierte Typen von Körperbewegungen, mit deren Hilfe diese Handlungen ausgeführt werden können. Wenn hinter meinem Rücken ein Bär auftaucht, muss ich mich umdrehen und – sofern ich ein geeignetes Messer habe – versuchen zuzustechen. Wegen der Äquivalenz zwischen akteurzentrierten und Repräsentationen mit dem Namen ‚Objekt-100‘ erben viele Repräsentationen mit diesem Namen diese beide Merkmale akteurzentrierter Repräsentationen. Mit anderen Worten: Repräsentationen mit dem Namen ‚Objekt-100‘ haben – zumindest häufig – dieselbe unmittelbare Handlungswirksamkeit wie die entsprechenden akteurzentrierten Repräsentationen. Und sie sind in derselben unmittelbaren Weise mit bestimmten Typen von Körperbewegungen verbunden, durch die AL bestimmte Handlungen ausführen kann.

Dies ist entscheidend, da für *de se* Einstellungen ebenfalls charakteristisch ist, dass sie eine spezifische kognitive Rolle und eine spezifische kausale Rolle im Hinblick auf die Handlungen einer Person spielen. Zumindest in vielen Fällen führen sie zu typischen egozentrischen Reaktionen dieser Person. Nehmen wir noch einmal Perrys Zuckerbeispiel. Bevor er bemerkte, dass er selbst den ganzen Schlammassel verursacht hatte, brachte ihn seine Überzeugung dazu, die Person mit der beschädigten Packung Zucker im Einkaufswagen zu suchen, um sie darauf hinzuweisen, welche Unordnung sie verursache. Danach änderte sich sein Verhalten auf typische Weise. Perry hörte auf, nach der anderen Person zu suchen und versuchte stattdessen, die beschädigten Packung Zucker in *seinem eigenen* Einkaufswagen in Ordnung zu bringen. Oder nehmen wir den Fall Ernst Machs, über den dieser selbst schreibt:

Ich stieg einmal nach einer anstrengenden nächtlichen Eisenbahnfahrt sehr ermüdet in einen Omnibus, eben als von der anderen Seite auch ein Mann hereinkam. „Was steigt doch da für ein herabgekommener Schulmeister ein“, dachte ich. (Mach 1911, 3 Fn. 1)

Dieser Gedanke wird bei Mach zu kaum mehr geführt haben als vielleicht zu einer gewissen Geringschätzung oder im besten Fall zu Mitleid. Doch plötzlich



stellte Mach fest, dass der Mann am anderen Ende des Busses niemand anderes war als er selbst.

Ich war es selbst, denn mir gegenüber befand sich ein großer Spiegel. Der Klassenhabitus war mir also viel geläufiger als mein Specialhabitus. (ebd.)

Diese Erkenntnis änderte die Dinge erheblich. Denn als er bemerkte, dass der Mann am anderen Ende des Busses niemand anderes als er selbst war, wird Mach wohl gedacht haben: *Ich* sehe aus wie herabgekommener Schulmeister. Und dieser Gedanke hatte ganz andere Konsequenzen als der erste – vielleicht begann Mach sich zu schämen oder zu versuchen, seine Kleider ein bisschen zu säubern.

Die Antwort auf die Frage, welche spezifischen Merkmale der computationalen Rolle des Namens ‚Objekt-100‘ dafür verantwortlich sind, dass dieser Name eine EGO-Art des Gegebenseins verkörpert, lautet also: Repräsentationen, die diesen Namen enthalten, spielen genau die kausale Rolle, die für *de se* Einstellungen charakteristisch ist – für Einstellungen, deren Inhalt angemessen nur unter Verwendung des Wortes ‚ich‘ wiedergegeben werden kann.

## 5.

Kognitive Systeme sind Wesen, die versuchen, Wissen über ihre Umwelt zu erwerben, um in dieser Umwelt besser zurechtzukommen. Zum effektiven Handeln benötigen sie vielfach jedoch nicht nur Wissen über ihre Umgebung, sondern auch Wissen über sich selbst. Jedes kognitive System muss wissen, ob *es* bedroht sind, wo *es* sich befinden, welche *seiner* Glieder einsatzfähig sind, was *es* braucht (Wasser, Energie, Ruhe), usw. In den meisten Fällen reicht es allerdings aus, diese Art von Selbstkenntnis in akteurzentrierten Repräsentationen zu speichern – Repräsentationen, in denen es zwar auch um das jeweilige kognitive System geht, in denen jedoch nicht explizit auf dieses System Bezug genommen wird. Repräsentationen mit expliziter Selbstbezugnahme werden für ein kognitives System erst unabdingbar, wenn es beginnt, Objekte in seiner Umwelt als Wesen zu repräsentieren, die ihrerseits kognitive Systeme sind, d.h. die ihrerseits über Repräsentationen ihrer Umwelt verfügen. Denn in den Repräsentationen seiner Mitwesen kommt das kognitive System selbst als Objekt vor – als etwas, von dem die Repräsentationen der Mitwesen handeln. Explizite Selbstkenntnis erwächst also aus der Erkenntnis eines kognitiven Systems, dass die anderen es genau so repräsentieren, wie es selbst die anderen repräsentiert. Spätestens wenn es repräsentieren will, welche Repräsentationen andere Wesen über es selbst haben, braucht ein kognitives System einen internen Namen für sich selbst; und damit wird es sich selbst zum möglichen Objekt der Repräsentation. Der letzte Schritt zu genuiner Selbstkenntnis besteht schließlich darin, dass das System beginnt, einen Zusammenhang zwischen den akteurzentrierten Repräsentationen, in denen es nur implizit um es selbst geht, und den neuen expliziten Selbstreprä-

sentationen herzustellen. Auf diese Weise erhalten diese neuen Repräsentationen genau die Rolle, die genuine *de se* Einstellungen charakterisiert.

## Literatur

- Beckermann, A. „Self-Consciousness in Cognitive Systems“. In: Ch. Kanzian, J. Quitterer, E. Runggaldier (Hg.) *Persons. An Interdisciplinary Approach*. Wien: öbv&hpt 2003, 174–188
- Descartes, R. *Meditationes de prima philosophia. Oeuvres des Descartes VII*. Publiées par C. Adam et P. Tannery. Nouvelle Présentation. Paris, J. Vrin, 1964–1976.
- Descartes, R. *Selected Philosophical Writings*, transl. by J. Cottingham, R. Stoothoff and D. Murdoch. Cambridge: Cambridge University Press 1988.
- Descartes, R. *The Philosophical Works of Descartes, Vol. 1.*, transl. by E.S. Haldane and G.R.T. Ross. Cambridge: Cambridge University Press 1911.
- Locke, J. *An Essay Concerning Human Understanding*, ed. by P.H. Nidditch, Oxford: Clarendon Press 1975.
- Lowe, E.J. 2000 *An Introduction to the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Mach, E. 1911 *Die Analyse der Empfindungen und das Verhältnis des Physischen zum Psychischen*, 6. Aufl. (1. Aufl. 1885), Jena.
- Perry, J. 1979 „The Problem of the Essential Indexical“, in J. Perry, *The Problem of the Essential Indexical and Other Essays*, Oxford: Oxford University Press 1993, 33–52.
- Perry, J. 1998 „Myself and I“, in Marcelo Stamm (Hg.) *Philosophie in synthetischer Absicht (Festschrift für Dieter Henrich)*, Stuttgart: Klett-Cotta, 83–103.
- Rosenberg, J.F. 1986 *The Thinking Self*, Philadelphia: Temple University Press.