

# Bachelorarbeit

in der Fakultät für Geschichtswissenschaft, Philosophie und Theologie  
der Universität Bielefeld  
zum Thema

## Funktionalismus, Paritätsprinzip und die These des erweiterten Geistes

vorgelegt von Fabian Hundertmark  
geboren am 24.03.1985 in Hameln

Betreuer:  
PD Dr. Holger Lyre und Dr. Peter Schulte

Bielefeld, 12. Mai 2009

### **Zusammenfassung**

Im folgenden Text werde ich mich mit dem Verhältnis von Funktionalismus und Paritätsprinzip befassen und die Frage klären, ob die Kombination beider Thesen dazu führt, dass es in unserer Welt externe geistige Zustände tatsächlich gibt. Zu diesem Zweck stelle ich zunächst den Funktionalismus und das Paritätsprinzip dar und kläre die Frage, ob die eine Theorie aus der anderen folgt. Da dies nicht der Fall ist, ergibt sich aus der Kombination beider Thesen der erweiterte Funktionalismus. Dieser ist wesentlicher Bestandteil einer Argumentation für die These des erweiterten Geistes. Ich werde jedoch zeigen, dass eine solche nicht möglich ist, da auf keine funktionalistische Theorie zurückgegriffen werden kann. Zum Schluss werde ich beweisen, dass ein auf dem erweiterten Funktionalismus basierendes Argument gegen ein in der Debatte viel besprochenes Beispiel nicht gültig ist, da es auf einer unplausiblen Annahme beruht.

# Inhaltsverzeichnis

<b>1</b>	<b>Der Funktionalismus</b>	<b>3</b>
1.1	Funktionen . . . . .	3
1.2	Funktionale Definitionen . . . . .	4
1.3	Funktionale Zustände und Prozesse in einem System . . . . .	5
1.4	Realisierung . . . . .	7
1.5	Geist als Funktion . . . . .	9
1.6	Rollen- und Realisiererfunktionalismus . . . . .	10
<b>2</b>	<b>Das Paritätsprinzip</b>	<b>11</b>
2.1	Geist und Kognition . . . . .	11
2.2	Die Lokalisierungsthese des Internalismus . . . . .	12
2.3	Das Paritätsprinzip . . . . .	13
<b>3</b>	<b>Das Verhältnis von Funktionalismus und Paritätsprinzip</b>	<b>15</b>
3.1	Folgt das Paritätsprinzip aus dem Funktionalismus? . . . . .	15
3.2	Folgt der Funktionalismus aus dem Paritätsprinzip? . . . . .	17
3.3	Sind Funktionalismus und Paritätsprinzip vereinbar? . . . . .	20
3.4	Der erweiterte Funktionalismus . . . . .	21
<b>4</b>	<b>Ein funktionales Argument für den faktisch erweiterten Geist</b>	<b>24</b>
4.1	Das Argument der kognitiven Äquivalenz . . . . .	25
4.2	Das Argument der funktionalen Äquivalenz . . . . .	26
4.3	Ein Argument auf Basis des funktionalen Paritätsprinzips . . . . .	26
4.4	Das Beispiel von Otto und seinem Notizbuch . . . . .	27
4.5	Das Problem des Fehlens einer funktionalistischen Theorie . . . . .	30
<b>5</b>	<b>Ein Argument gegen (Realisierer von) Überzeugungen in Ottos Notizbuch</b>	<b>30</b>
5.1	Der Aufbau der Argumentation . . . . .	31
5.2	Mögliche Wahrmacher der Prämissen 1 und 2 . . . . .	32
5.3	Die Prämisse 3 und der Psychofunktionalismus . . . . .	35
5.4	Ein Einwand gegen Prämisse 3 und den Psychofunktionalismus . . . . .	36
5.5	Der Common-Sense Funktionalismus die neue Prämisse 3 . . . . .	37
<b>6</b>	<b>Eine Rekapitulation</b>	<b>38</b>
	<b>Danksagungen</b>	<b>40</b>
	<b>Literatur</b>	<b>41</b>
	<b>Abbildungsverzeichnis</b>	<b>42</b>

# 1 Der Funktionalismus

Wenn wir im Alltag über den Menschen und sein Verhalten sprechen, so können wir dies auf verschiedene Art und Weise tun. Dies lässt sich an folgendem Beispiel zeigen.

**Ingas Museumsbesuch:** Inga kommt an einem Plakat vorbei, auf dem steht, dass im Museum eine Kunstausstellung mit Werken Yoko Onos zu sehen ist. Eine Stunde später ist Inga im Museum.

Eine Erklärung dieser Geschichte kann auf mindestens zwei verschiedene Arten erfolgen.

**Mentale Erklärung für Ingas Museumsbesuch:** Inga sieht das Plakat. Da sie weiß, welches Datum gerade ist, kommt sie zu dem Schluss, dass die Ausstellung am gleichen Tag stattfindet. Sie wollte schon immer einmal eine Ausstellung von Yoko Ono sehen und hat auch am heutigen Tag nichts anderes vor. Also überlegt sie kurz, wo sich das Museum befindet und geht hin.

**Physikalische Erklärung für Ingas Museumsbesuch:** Sonne strahlt auf das Plakat, das somit Licht zurückwirft. Dieses trifft auf Ingas Netzhaut und wird dort in elektrische Signale umgewandelt. Diese Signale erreichen das Gehirn und sorgen dort, zusammen mit anderen Informationen in Form bestimmter elektrochemischer Zustände des Gehirns, dafür, dass am Ende dieses Vorgangs Muskeln in einer bestimmten Art und Weise aktiviert werden. Durch diese Aktivierung bewegt sich Inga in die Richtung des Museums. Während dieser Bewegung sorgt das Gehirn durch ausgleichende Steuerung der Muskeln dafür, dass Inga nicht hinfällt oder vom Weg abkommt.

Es soll hier nicht um die Details gehen, sondern darum, dass uns beide Erklärungen richtig erscheinen. Doch haben wir es hier aller Wahrscheinlichkeit nach nicht mit einer doppelten Verursachung zu tun. Eine solche würde zum Beispiel vorliegen, wenn eine Glasscheibe genau im gleichen Moment von einem Stein getroffen wird, in dem sie auch durch Überdruck platzt. In einem solchen Fall wäre es tatsächlich unmöglich, eine einzelne oder mehrere zusammenwirkende Ursachen für das Kaputtgehen der Scheibe festzustellen.

Es wird also eine Theorie benötigt, die das Verhältnis des Mentalen zum Physikalischen erklären kann. Genau dies versucht der Funktionalismus.

## 1.1 Funktionen

Was ist eine Funktion?

Hier einige Beispiele: Wir reden im Alltag davon, dass ein Telefon nicht funktioniert. Biologen schreiben dem Herz die Funktion zu, Blut zu pumpen, oder behaupten, dass einzelne Zellen verschiedene Funktionen übernehmen können. Mathematisch gesehen ist eine Funktion eine Beziehung zwischen zwei Mengen, die jedem Element der einen Menge ein Element der anderen Menge zuordnet. So weist zum Beispiel die Funktion

$f(x) = x^2$  der Zahl 3 die Zahl 9 zu. Einen ähnlichen Funktionsbegriff finden wir auch in der Informatik, in der eine Funktion eine Unterroutine eines Programms darstellt, die bestimmte Eingabeparameter bekommt und bestimmte Ausgabeparameter zurückgibt. Nun habe ich einige Beispiele für Funktionen in unterschiedlichen Bereichen gegeben. Auf alle trifft folgender Satz zu:

Funktionen erzeugen bei den Eingangszuständen  $E_1, \dots, E_n$  die Ausgangszustände  $A_1, \dots, A_n$ .

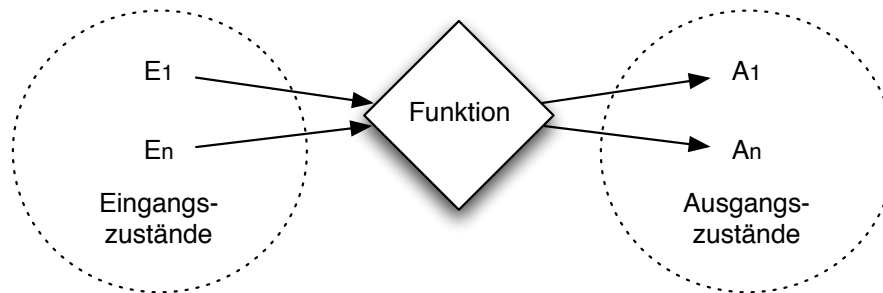


Abbildung 1: Eine Funktion mit den Eingangszuständen  $E_1$  und  $E_n$  und den Ausgangszuständen  $A_1$  und  $A_n$ .

So besteht zum Beispiel die Funktion eines Telefons darin, akustische Signale in bestimmte elektrische Signale umzuwandeln und umgekehrt. Eine Funktion in der Mathematik oder Informatik erhält Zahlen und gibt Zahlen zurück. Betrachten wir die Funktion  $f(x) = x^2$ , so liefert sie zum Beispiel für die Eingangszustände 1 ( $E_1$ ) und 2 ( $E_2$ ) die Ausgangszustände 1 ( $A_1$ ) und 4 ( $A_2$ ) zurück. Zwar brauchen wir eine gewisse Zeit, um dies zu berechnen, dennoch würden wir nicht davon sprechen, dass mathematische Funktionen eine bestimmte Zeit brauchen, um ihre Arbeit zu verrichten. Folglich muss zwischen dem Vorliegen der Eingangszustände und dem Ausgeben der Ausgangszustände keine Zeit vergehen, damit es sich um eine Funktion handelt.

## 1.2 Funktionale Definitionen

Wenn wir die Frage stellen, was ein Ding zu einem Ding einer bestimmten Art macht, so fallen die Antworten je nach Gegenstandsbereich relativ unterschiedlich aus.

**Eine bestimmte Zusammensetzung als Definiens**<sup>1</sup>: Etwas ist Wasser, wenn es die Mikrostruktur  $H_2O$  besitzt.

**Eine bestimmte evolutionäre Geschichte als Definiens** : Etwas ist ein Krokodil, wenn es eine bestimmte Evolutionsgeschichte hat.

Doch finden wir auch Definitionen, die mit der Funktion eines Dings zu tun haben. So ist etwas zum Beispiel eine Pumpe, wenn es einer Flüssigkeit oder einem Gas kinetische

<sup>1</sup>Das Definiens ist das, wodurch ein Begriff definiert wird.

Energie oder potentielle Energie in Form von Druck zuführt.

Dabei scheint es zunächst egal zu sein, welche Zusammensetzung oder Struktur das entsprechende Ding hat. So kann eine Pumpe aus Gold, Stein, Eis oder – im Fall des Herzens – aus Zellen bestehen. Auch können verschiedene Pumpen *im Detail* anders funktionieren. So handelt es sich bei Schöpfwerken (einfachen Hebevorrichtungen für Flüssigkeiten), Blasebälgen oder auch bei den in Dampflokomotiven eingesetzten Kolbenpumpen um völlig verschiedene Konstruktionen.

Ebenso wenig entscheidend scheint auf den ersten Blick die Geschichte des Dings zu sein. Ob die Pumpe von Menschen oder Außerirdischen hergestellt ist, ob sie durch Evolution oder puren Zufall entstanden ist, ist zunächst egal.<sup>2</sup> Rein funktionale Definitionen haben folgende Form:

**Rein funktionale Definitionen::**  $X$  ist ein Ding vom Typ  $T$ , wenn  $X$  aus den Eingangszuständen  $E_1, \dots, E_n$  die Ausgangszustände  $A_1, \dots, A_n$  erzeugt.

### 1.3 Funktionale Zustände und Prozesse in einem System

Ohne an dieser Stelle allzu sehr auf die Systemtheorie eingehen zu wollen, werde ich kurz versuchen, das Verhältnis von Funktionen zu Systemen klar zu machen. Zu diesem Zweck werde ich mich mit folgendem Beispiel beschäftigen<sup>3</sup>:

**Der Colaautomat:** Wenn wir einen Getränkeautomaten vor uns haben, der nur Fünfzigcent- und Ein-Euro-Strücke aufnimmt, der nur Coladosen ausgibt und bei dem eine Cola einen Euro kostet, dann sind die möglichen Eingangszustände Fünfzig-Cent-Stücke und Ein-Euro-Strücke im Münzeinwurfslitz und die möglichen Ausgangszustände eine Cola im Colaausgabeschacht, sowie ein Fünfzig-Cent-Stück im Münzausgabeschacht. Doch wie kann man beschreiben, dass eine Cola einen Euro kostet?

Es reicht offensichtlich nicht aus, zu sagen, dass der Automat einfach die Funktion hat, eine Cola auszuwerfen, wenn man einen Euro hinein tut, da so nicht erklärt wäre, warum er manchmal – nämlich wenn schon ein Fünfzig-Cent-Stück eingeworfen wurde – auch ein Fünfzig-Cent-Stück und eine Cola auswirft oder warum er manchmal auch eine Cola auswirft, wenn man nur fünfzig Cent einwirft.<sup>4</sup>

Folglich müssen wir annehmen, dass irgendetwas im System anzeigt, ob schon fünfzig Cent eingeworfen wurden. Dies geht jedoch schon über normale Funktionen ohne interne Zustände hinaus. Der Colaautomat kann also Systemzustände haben, die durch Funktionen verändert werden können.<sup>5</sup>

---

<sup>2</sup>Zugegebenermaßen müssen eventuell doch geschichtliche Informationen einbezogen werden, wenn es zum Beispiel um die Frage geht, ob etwas noch eine Pumpe ist, wenn es seine Funktion nicht mehr erfüllen kann.

<sup>3</sup>Das Beispiel stammt aus Ned Blocks Aufsätzen „Troubles with functionalism“ und „What is functionalism?“ und wurde von Beckermann (2000, S. 143f.) aufgenommen.

<sup>4</sup>Jemand, der schon weiß worauf diese Erklärung hinauslaufen soll, kann erkennen, dass es sich hier um den selben Schwachpunkt wie beim Behaviorismus handelt.

<sup>5</sup>Auf Abbildung 2 finden wir eine schematische Darstellung eines solchen Colaautomaten.

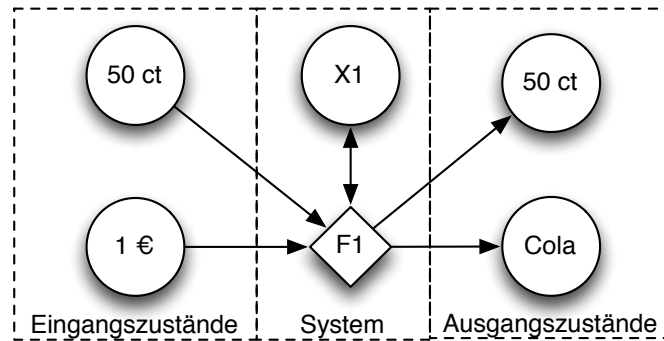


Abbildung 2: Der schematische Aufbau des Colaautomatensystems.

So hat die Funktion F1 des Getränkeautomaten vier mögliche Eingangszustände:

**A1** Einwurf von fünfzig Cent im Systemzustand 1 (S1)

**A2** Einwurf eines Euros im Systemzustand 1 (S1)

**A3** Einwurf von fünfzig Cent im Systemzustand 2 (S2)

**A4** Einwurf eines Euros im Systemzustand 2 (S2)

Zudem gibt es 8 mögliche Ausgangszustände:

**E1** Systemzustand 1 (S1)

**E2** Systemzustand 2 (S2)

**E3** Fünfzig Cent im Münzausgabeschacht und Systemzustand 1 (S1)

**E4** Fünfzig Cent im Münzausgabeschacht und Systemzustand 2 (S2)

**E5** Eine Coladose im Getränkeauswurfschacht und Systemzustand 1 (S1)

**E6** Eine Coladose im Getränkeauswurfschacht und Systemzustand 2 (S2)

**E7** Fünfzig Cent im Münzausgabeschacht, eine Coladose im Getränkeauswurfschacht und Systemzustand 1 (S1)

**E8** Fünfzig Cent im Münzausgabeschacht, eine Coladose im Getränkeauswurfschacht und Systemzustand 2 (S2)

Wie wir sehen, erhöht die Möglichkeit zweier Systemzustände, die sowohl Eingabe- als auch Ausgabezustände einer Funktion sein können, die Komplexität eines Systems erheblich.

Da die Prozesse im Colaautomaten bei jedem möglichen Eingangszustand nur zu einem möglichen Ausgangszustand führen und da ein Ausgangszustand von zwei Eingangszuständen erreicht werden kann, werden in der Praxis nur die Endzustände E2, E5 und E7 tatsächlich vom Automaten erzeugt.<sup>6</sup>

<sup>6</sup>Vorausgesetzt, dass sich Anfangs nie eine Coladose oder ein Fünfzig-Cent-Stück im entsprechenden Schacht befinden.

**P1** A1 führt zu E2

**P2** A2 führt zu E5

**P3** A3 führt zu E5

**P4** A4 führt zu E7

Die Information, ob schon ein Fünfzig-Cent-Stück eingeworfen wurde, wird durch einem Systemzustand gespeichert, der im Colaautomatensystem möglich sein muss. Dies wissen wir alleine durch Beobachtung des Systemverhaltens und ohne in den Automaten schauen zu müssen. Wir wissen zunächst natürlich nicht, wie der Systemzustand S2 in diesem Automaten umgesetzt ist. Einen solchen Zustand, der alleine durch die Wirkung auf die Funktion eines Systems bestimmt ist, nennt man einen „funktionalen Zustand“. Beckermann schreibt zu diesen:

**Funktionale Zustände:** „Funktionale Zustände sind Zustände eines Systems, die allein durch ihre *kausale Rolle* charakterisiert sind – d.h. durch die Ereignisse außerhalb des Systems, durch die sie verursacht werden (*inputs*), durch das, was sie selbst außerhalb des Systems verursachen (*output*), und durch ihre *kausalen Relationen zu anderen Systemzuständen* derselben Art.“(Beckermann, 2000, S. 142, Hervorhebungen im Original)

Dem entsprechend wären funktionale Prozesse wie folgt zu beschreiben:

**Funktionale Prozesse:** Funktionale Prozesse sind Prozesse eines Systems, die dadurch charakterisiert sind, durch welche Systeminputs und/oder Systemzustände sie verursacht werden und zu welchen Outputs und/oder Systemzuständen sie führen.

## 1.4 Realisierung

Ebenso wie es Begriffe gibt, die alleine durch ihre Funktion charakterisiert sind, sind bestimmte Systemtypen durch ihren funktionalen Aufbau und somit auch durch ihre möglichen funktionalen Zustände und Prozesse definiert.

**Funktionale Definition eines Systems:** Ein System S ist vom Systemtyp ST, wenn in S bestimmte – für ST nötige – Funktionen realisiert sind, die auf eine bestimmte – für ST nötige – Art und Weise mit Inputs, Outputs und systeminternen Zuständen gekoppelt sind.

Was lässt sich unter „Realisierung“ verstehen?

**Zur Realisierung funktionaler Zustände** schreibt Beckermann:

„Wenn ein System  $S$  die funktionalen Zustände  $Z_1, \dots, Z_n$  annehmen kann, dann werden diese Zustände genau dann durch die physikalischen Zustände  $P_1, \dots, P_n$  von  $S$  realisiert, wenn diese physischen Zustände genau die kausalen Rollen innehaben, durch die die funktionalen Zustände  $Z_1, \dots, Z_n$  charakterisiert sind.“ (Beckermann, 2000, S. 148)

Im Colaautomaten gibt es zwei funktionale Systemzustände, wenn wir Input- und Outputzustände nicht mit einbeziehen. Um die Frage zu beantworten, welcher Systemzustand auf welche Weise realisiert wird, müssen wir schauen, welche Zustände die relevanten kausalen Rollen innehaben. Wir müssen folglich fragen, was im Automaten dafür sorgt, dass die Prozesse  $P_1$  oder  $P_2$  ausgelöst werden, und was dazu führt, dass  $P_3$  oder  $P_4$  ausgelöst werden.

**Ein elektrisches System:** Machen wir den Colaautomaten auf und schauen hinein, so wäre es zum Beispiel möglich, dass der Automat einen Sensor hat, der Fünfzig-Cent-Stücke anhand der magnetischen Merkmale von Münzen unterscheiden kann. Immer wenn auf diese Weise ein Fünfzig-Cent-Stück erkannt wird, wird ein Stromstoßschalter (ein elektromagnetisch umlegbarer Schalter) umgelegt. Dieser Schalter sorgt in einer Stellung dafür, dass die Klappe, die normalerweise die Coladose zurückhält, dann und nur dann geöffnet wird, wenn der Münzerkenner das Signal gibt, das anzeigt, dass ein Euro eingeworfen wurde. Im anderen Schalterzustand wird die Klappe auch vom Signal ausgelöst, das beim Einwurf von Fünfzig-Cent-Stücken entsteht. Zusätzlich wird beim Einwurf eines Euros auch eine Klappe geöffnet, die ein Fünfzig-Cent-Stück davon abhält, in den Geldausgabeschacht zu fallen. In diesem Fall wäre der Systemzustand  $S_1$  durch die eine Schalterstellung und Systemzustand  $S_2$  durch eine andere Schalterstellung realisiert.

**Ein biologisches System:** Dieselbe Funktionalität wäre jedoch auch gewährleistet, wenn im Colaautomaten zwei Zwerge sitzen würden. Der eine Zwerg sitzt unter dem Münzeinwurfschlitz auf einem Hocker. Wird eine Münze eingeworfen, so schaut sich der Zwerg diese an. Ist es ein Fünfzig-Cent-Stück und liegt seine Hand noch nicht auf der Schulter des anderen Zwerges, legt er sie auf diese. Liegt sie hingegen schon dort, kneift er den anderen Zwerg und nimmt dann seine Hand von der Schulter herunter. Erkennt er, dass die eingeworfene Münze ein Ein-Euro-Strück ist und liegt seine Hand auf der Schulter des anderen Zwerges, kneift er diesen, wirft eine Fünfzigcentmünze in den Münz- ausgabeschacht und nimmt die Hand von der Schulter. Identifiziert er das eingeworfene Geldstück als Ein-Euro-Strück und hat er die Hand nicht auf der Schulter des anderen Zwerges, kneift er diesen nur. Immer wenn der andere Zwerg gekniffen wird, legt er eine Coladose in den Ausgabeschacht. In diesem Colaautomaten wäre der Systemzustand  $S_2$  durch die Hand des einen Zwerges auf der Schulter des anderen Zwerges realisiert. Der Systemzustand  $S_1$  hingegen wäre dadurch realisiert, dass die Hand des Zwerges unter dem Münzeinwurfschacht gerade nicht auf der Schulter des anderen Zwerges liegt.



**multiple Realisierbarkeit:** Die Systemzustände S1 und S2 sind in beiden Systemen unterschiedlich realisiert. Wäre der Systemtyp des Colaautomaten rein funktional definiert, wäre sowohl das elektrische als auch das biologische System ein System vom Typ „Colaautomat“. Der selbe Systemtyp könnte natürlich ebenso gut durch rein mechanische, hydraulische oder sonstige Vorgänge und Zustände realisiert werden. Man kann also sagen: „Alle Systeme, die [...] funktional [...] charakterisiert sind, sind ebenso multirealisierbar wie die funktionalen Zustände, die sie annehmen können.“(Beckermann, 2000, S. 152)

## 1.5 Geist als Funktion

Kommen wir zum Thema der mentalen Zustände. Stellen Sie sich folgende Situation vor: Sie begegnen einem Mann im ICE von Bielefeld nach Hannover. Sie kommen mit ihm ins Gespräch und er erzählt ihnen, dass er dort eine Frau treffen will, die er vor ein paar Monaten kennengelernt hat. Er schwärmt von dieser Frau und erzählt eine halbe Stunde lang, wie schön, witzig und intelligent sie ist und wie glücklich er ist, sie zu kennen. Sie kommen zu der Überzeugung, dass er sich im Moment nichts sehnlicher wünscht, als diese Frau in Hannover zu treffen und dass er voller Vorfreude ist. Kurz bevor Sie in Hannover aussteigen, erzählt Ihnen der Mann, dass er vor ein paar Jahren eine starke Verletzung am Gehirn hatte. Inzwischen sei er jedoch vollständig genesen. Die beschädigten Hirnteile könne er zwar nicht mehr nutzen, dafür hätten aber andere Teile die Arbeit dieser übernommen. Andere Teile seiner Gehirnfunktionen seien durch eine neuartige Technologie, auf Basis sehr kleiner Microchips, ersetzt worden.

Vorausgesetzt Sie glauben ihm: Würde die Information über seine Hirnschädigung und die Behandlung dazu führen, dass Sie ihre Überzeugungen bezüglich der Gefühle und Wünsche des Mannes die Frau betreffend ändern würden?

Ich würde es nicht und glaube auch, dass es die meisten Menschen nicht tun würden. Es scheint nicht entscheidend für geistige Zustände – die Überzeugungen und Wünsche des Mannes – zu sein, in welchem Zustand das Gehirn der Person genau ist, die diese Zustände hat. Wie lässt sich dies erklären? Ich sehe hier drei Möglichkeiten:

1. Man behauptet, dass unsere Intuitionen irreführend sind und dass der Mann in Wirklichkeit keine geistigen Zustände hat oder zumindest notwendigerweise andere haben muss, als Menschen mit einem normalen Gehirn.
2. Man erkennt die Intuition an, leugnet aber, dass unsere geistigen Zustände durch Gehirnzustände bestimmt werden.
3. Man erkennt die Intuition an und behauptet, dass geistige Zustände multipel realisierbar sind.

**Möglichkeit 1** ist wenig attraktiv, da wir diesem Argumentationsmuster folgend nicht nur diesem Mann geistige Zustände absprechen müssen, sondern allen Menschen, deren

Gehirn etwas anders aufgebaut ist. Wir müssten quasi eine ausführliche Untersuchung über die Substanz des Gehirns und über die an bestimmten Aufgaben beteiligten Areale durchführen, um tatsächlich entscheiden zu können, ob ein Wesen Wünsche hat oder nicht. Dies scheint maßlos übertrieben und würde für die meisten von uns bedeuten, dass wir nicht einmal wissen können, ob wir selbst geistige Zustände haben, da unser Gehirn möglicherweise etwas anders aufgebaut ist als andere.

**Möglichkeit 2** scheint unseren wissenschaftlichen Erkenntnissen zu widersprechen, denen zufolge unsere Handlungen, Entscheidungen und Wahrnehmungen maßgeblich mit unserem Gehirn zu tun haben.

**Möglichkeit 3** hingegen scheint plausibel zu sein, da sie erlaubt, dass *unsere Gehirnzustände unsere geistigen Zustände* bestimmen, während *die geistigen Zustände des Mannes im Zug sowohl von seinem Gehirn, als auch den Microchips in seinem Kopf* abhängen. Ebenso können wir Außerirdischen aus Silicium oder Robotern mit pneumatischen Gehirnen prinzipiell geistige Zustände zuschreiben.

Eine Möglichkeit multiple Realisierbarkeit von Systemen zu gewährleisten ist es, die Systeme und ihre Zustände funktional zu definieren. Hier kommt der Funktionalismus ins Spiel, dessen zentrale These folgende ist:

**Funktionalismus:** „Mentale Zustände sind ihrer Natur nach *funktionale Zustände*.“ (Beckermann, 2000, S. 142, Hervorhebung im Original)

oder wie es Shapiro ausdrückt:

„Real minds, according to the functionalists, are nothing more than the realizations of functionally defined mental states.“ (Shapiro, 2008, S. 7)

Ein System hat also genau dann eine bestimmte Überzeugung, wenn es sich in einem Zustand befindet, der durch bestimmte Eingangszustände oder bestimmte systeminterne Zustände verursacht wird oder der bestimmte Ausgangszustände oder Systemzustände verursacht.<sup>7</sup>

## 1.6 Rollen- und Realisiererfunktionalismus

Eine These, die wohl kein Funktionalist bestreiten würde, ist folgende:

**minimaler Funktionalismus:** Für alle mentalen Zustände M und alle Systeme S gilt: S ist genau dann im mentalen Zustand M, wenn es sich im funktionalen Zustand Z befindet.

---

<sup>7</sup>Wobei alle „...oder...“ in diesem Satz als „...und/oder...“ zu verstehen sind und nicht als „entweder...oder...“.

Die Meinungen gehen jedoch bei der Frage auseinander, welche der folgenden Thesen zudem akzeptiert werden sollte (vgl. Braddon-Mitchell und Jackson, 1996 S. 100f., Beckermann, 2000 S. 142 und Levin, 2009):

**Rollenfunktionalismus:** Minimaler Funktionalismus und  $M = Z$

**Realisiererfunktionalismus:** Minimaler Funktionalismus und  $M$  ist das, was im System  $S$  die Rolle  $Z$  erfüllt.

Es geht hier darum, ob zum Beispiel Schmerz nur eine Eigenschaft eines Systems ist oder ob es sich um ein bestimmtes „Ding“ in der Welt – wie zum Beispiel einen Gehirnzustand – handelt. Dem Realisiererfunktionalismus zufolge könnte man mentale Zustände lokalisieren und so zum Beispiel behaupten, dass sich ein Wunsch oder eine Überzeugung an einer bestimmten Stelle im Gehirn befindet. Der Rollenfunktionalismus hingegen würde nur dem ganzen System die Eigenschaft zuschreiben, Schmerzen zu haben.

Ich will nicht die Argumente für die eine oder andere Position abwägen, sondern beide im Blick behalten.

## 2 Das Paritätsprinzip

In diesem Abschnitt soll es um das Paritätsprinzip gehen. Dieses stellt eine These über den Ort – beziehungsweise über den Ort der Realisierung – kognitiver Zustände und Prozesse dar. Doch zunächst will ich kurz auf das Verhältnis von Geist und Kognition sowie auf die klassische Auffassung eingehen, die durch das Paritätsprinzip herausgefordert wird.

### 2.1 Geist und Kognition

Der Funktionalismus, wie ich ihn hier vorgestellt habe, ist zunächst eine These über den Geist. Im Paritätsprinzip hingegen ist von „Kognition“ die Rede. Wenn dieser Text das Verhältnis beider Theorien zueinander klären soll, so scheint eine nahe liegende Frage zu sein, in welchem Verhältnis Geist und Kognition stehen.

Auf diese Frage sind verschiedene Antworten möglich. So kann man zum Einen, der Bedeutung des lateinischen „cognoscere“ – welches im Deutschen „erkennen“, „erfahren“ oder „kennen lernen“ heißt – folgend, kognitive Zustände als diejenigen mentalen Zustände ansehen, die die Welt korrekt oder inkorrekt repräsentieren. Dieser Definition zufolge, wären kognitive Zustände zum Beispiel Wahrnehmungen und Überzeugungen, jedoch nicht Wünsche.

Andererseits kann man die Begriffe „Geist“ und „Kognition“ auch synonym verwenden. Dem entsprechend würde gelten:

- geistige Zustände = kognitive Zustände

- geistige Prozesse = kognitive Prozesse
- geistige Fähigkeiten = kognitive Fähigkeiten

Auch wenn in der Debatte um den erweiterten Geist und ums Paritätsprinzip meist Beispiele kognitiver Prozesse und Zustände im oberen Sinne eine Rolle spielen, halte ich dies für sinnvoll, da die in der Psychologie traditionell nicht als kognitiv betrachteten Zustände und Vorgänge, wie Emotion oder Motivation, untrennbar mit Kognition verbunden und daher nicht klar von dieser abgrenzbar zu sein scheinen. (vgl. Strube u. a., 1995, S. 300).

Daher können wir sagen, dass sich der Funktionalismus, das Paritätsprinzip und der erweiterte Geist auf den gleichen Gegenstandsbereich beziehen.

## 2.2 Die Lokalisierungsthese des Internalismus

Die klassische Auffassung über die Lokalisierung des Mentalen soll hier exemplarisch an zwei Textausschnitten dargestellt werden. Horst M. Müller schreibt in einer Einführung über die neurobiologischen Grundlagen der Sprachfähigkeit:

„Da grundsätzlich alle psychischen Phänomene auf ein neuronales Substrat zurückgeführt werden können, hat auch die menschliche Sprachfähigkeit ihre Ursache in Prozessen, Funktionen und Eigenschaften des menschlichen Nervensystems.“ (Müller, 2003, S. 57)

Weiter unten findet sich folgender Satz:

„Eine stark schematisierte Darstellung der Vorgänge im ZNS, die letztlich die Kognition des jeweiligen Organismus ausmachen, zeigt Abbildung 4.2 [hier im Text Abbildung 3].“ (Müller, 2003, S. 58)

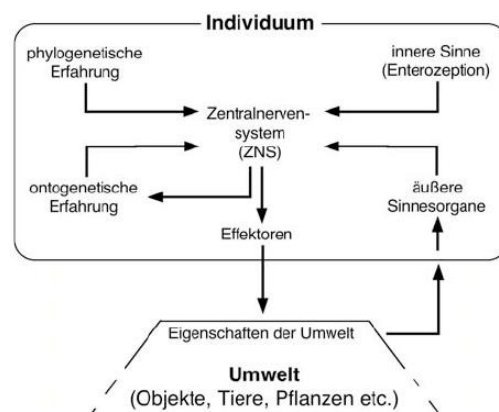


Abbildung 3: Individuum-Umwelt-Interaktion (aus Müller, 2003, S. 59)

Müller geht scheinbar davon aus, dass Kognition im zentralen Nervensystem *und nur im zentralen Nervensystem* stattfindet. Damit scheint er eine Extremposition der Auffassung

zu vertreten, die Mark Rowlands als die Lokalisierungsthese (im Original: „Location Claim“) des Internalismus bezeichnet:

**Die Lokalisierungsthese des Internalismus:** „any mental phenomenon is spatially located inside the boundaries of the subject, [SJ], that has or undergoes it.“(Rowlands, 2003, S. 13)

Dabei stellt normalerweise die Haut die Grenze dar, die für den Internalisten entscheidend ist. Während – nach Rowlands – der internalistischen Lokalisierungsthese zufolge mentale Phänomene und somit auch kognitive Zustände und Prozesse im Körper eines Wesens stattfinden, lokalisiert Müller diese, wie oben schon beschrieben, nur im zentralen Nervensystem.

Die Lokalisierungsthese des Internalismus lässt sich auf drei verschiedene Arten und Weisen vertreten:

**Analytisch:** Es gehört begrifflich zu geistigen Phänomenen, dass sie oder ihre Realisierungen nur in den Grenzen des Subjektes zu lokalisieren sind, dem sie zugesprochen werden.

**Nomologisch:** Unsere Naturgesetze erlauben es nur, dass geistige Phänomene oder ihre Realisierungen in den Grenzen des Subjektes zu lokalisieren sind, dem sie zugesprochen werden.

**Kontingent:** Geistige Phänomene oder ihre Realisierungen sind in unserer Welt in den Grenzen des Subjektes zu finden, dem sie zugesprochen werden.

Dabei lässt sich die nomologische Behauptung nicht bestreiten, ohne auch die analytische Behauptung abzulehnen. Auch schließt natürlich die Ablehnung der These, dass die Lokalisierungsthese in unserer Welt wahr ist, die logische und nomologische Möglichkeit einer solchen Ablehnung ein.

### 2.3 Das Paritätsprinzip

Dass das zentrale Nervensystem eine entscheidende Rolle für unsere Kognition spielt, wollen sicherlich wenige bestreiten. Dennoch ist in den letzten Jahren – vor allem durch Clark und Chalmers (1998) – eine Diskussion in den Kognitionswissenschaften und der Philosophie entfacht worden, ob Kognition *ausschließlich* im diesem stattfindet. Dabei nimmt das so genannte Paritätsprinzip eine entscheidende Rolle ein. Die ursprüngliche Formulierung dieses Prinzips findet sich in Clark und Chalmers (1998) und lautet wie folgt:

„If, as we confront some task, a part of the world functions as a process which, where it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process.“(Clark und Chalmers, 1998, S. 8)

Die Frage, die das Paritätsprinzip beantwortet, ist, wann ein Prozess in der Welt – also außerhalb des Körpers – als Teil eines kognitiven Prozesses anzusehen ist. Die Antwort, die das Paritätsprinzip gibt, ist einfach: Ein Prozess außerhalb des Körpers ist Teil eines kognitiven Prozesses, wenn er innerhalb des Körpers Teil eines kognitiven Prozesses wäre. Zehn Jahre später erläutert Clark das Paritätsprinzip wie folgt:

„[F]or the purposes of identifying the material vehicles of cognitive states and processes, we should (normatively speaking) ignore the old metabolic boundaries of skin and skull and attend to the computational and functional organization of the problem-solving whole.“(Clark, 2008, S. 77)

Clark erweitert in seiner Erläuterung das Paritätsprinzip auf kognitive Zustände und erlaubt auch, dass nicht nur Teilprozesse, sondern auch ganze Prozesse extern sein können. Zudem betont er, dass wir das Paritätsprinzip akzeptieren *sollten*. Es soll also nicht darum gehen, dass wir unsere Begriffe im Alltag oder in der Wissenschaft schon so verwenden, als würden wir das Paritätsprinzip akzeptieren. Viel mehr ist Clark zufolge entscheidend, dass eine solche Akzeptanz des Paritätsprinzips bessere Erklärungen ermöglicht (vgl. Clark und Chalmers, 1998, S. 14). Warum dies aber wünschenswert ist, ist letzten Endes eine normative Frage. Seine Erläuterung lässt sich in drei Teilthesen zerlegen:

1. Kognitive Zustände und Prozesse oder ihre Realisierungen sind nicht notwendigerweise nur innerhalb der Körpergrenzen zu lokalisieren.
2. Kognitive Zustände und Prozesse oder ihre Realisierungen sind durch die funktionale und computationale Organisation des problemlösenden Ganzen zu bestimmen.
3. These 1 und 2 sollten wir aus normativen Gründen akzeptieren.

Gerade die zweite Teilthese zeigt deutlich, dass Clark ein funktionalistisches Verständnis von Kognition hat. Adams und Aizawa (2008, S. 134) folgend werde ich jedoch eine Lesart des Paritätsprinzips bevorzugen, die nicht schon den Begriff „Funktionalismus“ enthält. Des Weiteren werde ich – der Originalformulierung folgend – das Paritätsprinzip nicht normativ lesen. Somit gehört auch die dritte Teilthese nicht zum Paritätsprinzip. Dennoch lässt sich natürlich sagen, dass wir es aus normativen Gründen akzeptieren sollten. Das Paritätsprinzip lautet dem entsprechend:

**Paritätsprinzip:** Wenn sich der Zustand Z oder der Prozess P teilweise oder komplett außerhalb der körperlichen Grenzen von Subjekt SJ befindet und Z oder P ansonsten im relevanten Sinne kognitiv äquivalent zu internen geistigen Zuständen oder Prozessen von SJ ist, ist Z oder P kognitiv und SJ zuzuschreiben.

Es soll also um kognitive und nicht zwangsläufig um funktionale Äquivalenz gehen. Dennoch muss kognitive Äquivalenz auf die eine oder andere Weise ausbuchstabiert werden. Tut man dies mit Hilfe des Funktionalismus, folgt – wie ich zeigen werde – auch

Clarks zweite These.

Im Bezug auf die Lokalisierungsthese des Internalismus lässt sich sagen, dass das Paritätsprinzip nur die logische Möglichkeit voraussetzt, Kognition auch außerhalb der Grenzen des Subjektes zu lokalisieren.

### 3 Das Verhältnis von Funktionalismus und Paritätsprinzip

Nach dieser Einführung und Definition der relevanten Begriffe soll es nun um das Verhältnis des Funktionalismus zum Paritätsprinzip gehen. Dabei werde ich drei Fragen beantworten:

1. Enthält der Funktionalismus das Paritätsprinzip?
2. Setzt das Paritätsprinzip bereits den Funktionalismus voraus?
3. Sind Paritätsprinzip und Funktionalismus vereinbar?

#### 3.1 Folgt das Paritätsprinzip aus dem Funktionalismus?

Zunächst soll es um die Überprüfung der folgenden These gehen:

Wenn der Funktionalismus wahr ist, dann ist auch das Paritätsprinzip wahr.

Ich werde mich im Folgenden dafür aussprechen, dass diese Behauptung nicht akzeptiert werden kann. Dabei werde ich folgende Argumentation verwenden:

**Prämisse 1:** Wenn es möglich ist, dass x und y wahr sind und wenn es nicht möglich ist, dass y und z wahr sind, dann folgt aus der Wahrheit von x nicht die Wahrheit von z.<sup>8</sup>

**Prämisse 2:** Es gibt eine Theorie T, für die gilt: Es ist möglich, dass T wahr ist und, dass der Funktionalismus wahr ist und es ist nicht möglich, dass T wahr ist und, dass das Paritätsprinzip wahr ist.

**Konklusion:** Aus der Wahrheit des Funktionalismus folgt nicht die Wahrheit des Paritätsprinzips.

Dieses Argument ist schlüssig. Ob es gültig ist, hängt von der Wahrheit der Prämissen ab. Daher werde ich nun einzeln für sie argumentieren.

---

<sup>8</sup>Hier können unter x, y und z sowohl einfache Propositionen als auch als komplexe Theorien verstanden werden.

**Prämisse 1:** Wenn es möglich ist, dass x und y wahr sind und wenn es nicht möglich ist, dass y und z wahr sind, dann folgt aus der Wahrheit von x nicht die Wahrheit von z.

Da ich glaube, dass diese Prämisse von selbst einleuchten sollte, werde ich hier nur ein Beispiel anführen. Setzen wir beispielsweise für die Variablen x, y und z folgende Propositionen ein,

**x** = „Karlo lebt im Wasser.“

**y** = „Karlo ist ein Delphin.“

**z** = „Karlo ist ein Fisch.“

, so ist es möglich, dass Karlo ein Delphin ist und im Wasser lebt. Es ist nicht möglich, dass Karlo ein Delphin und ein Fisch ist, da Delphine keine Fische sind. Nach Prämisse 1 folgt daraus, dass Karlo im Wasser lebt nicht, dass Karlo ein Fisch ist. Dies ist zweifellos richtig, da nicht nur Fische im Wasser leben.

**Prämisse 2:** Es gibt eine Theorie T, für die gilt: Es ist möglich, dass T wahr ist und, dass der Funktionalismus wahr ist und es ist nicht möglich, dass T wahr ist und, dass das Paritätsprinzip wahr ist.

Schauen wir uns noch einmal die These des Funktionalismus in Bezug auf funktionale Zustände an:

**minimaler Funktionalismus:** Für alle mentalen Zustände M und alle Systeme S gilt: S ist genau dann im mentalen Zustand M, wenn es sich im funktionalen Zustand Z befindet.

**Funktionale Zustände:** „Funktionale Zustände sind Zustände eines Systems, die allein durch ihre *kausale Rolle* charakterisiert sind – d.h. durch die Ereignisse außerhalb des Systems, durch die sie verursacht werden (*inputs*), durch das, was sie selbst außerhalb des Systems verursachen (*output*), und durch ihre *kausalen Relationen zu anderen Systemzuständen der selben Art.*“ (Beckermann, 2000, S. 142, Hervorhebungen im Original)

Wenn ein funktionaler Zustand durch die Inputs und Outputs eines Systems bestimmt ist, benötigt der Funktionalismus klarerweise eine Abgrenzung des Systems von der Umwelt. Über eine solche Abgrenzung hingegen herrscht Uneinigkeit, sodass unter anderem folgende Positionen möglich sind:

1. *Inputs* sind die elektrochemischen Signale, die das Gehirn von den Sinnesorganen erhält, und *outputs* die elektrochemischen Signale, die das Gehirn an die Muskeln schickt.



2. *Inputs* sind die physikalischen Reize, die von unseren Sinnesorganen verarbeitet werden [...] und *outputs* sind die Bewegungen unserer Gliedmaßen.
3. *Inputs* sind die verschiedenen Umweltsituationen, in denen wir uns befinden, und *outputs* sind die Veränderungen in unserer Umwelt, die wir durch unsere Handlungen hervorrufen.

Beckermann (2000, S. 175, Hervorhebungen im Original)

Dabei lassen sich funktionalistische Theorien, die die erste oder zweite Input- und Outputdefinition nutzen, als „short-arm“, Theorien, die die dritte Möglichkeit wählen hingegen als „long-arm“ Theorien bezeichnen (vgl. Levin, 2009).<sup>9</sup> Es spricht also, auch wenn man den Funktionalismus akzeptiert, prima facie nichts dagegen, eine funktionalistische „short-arm“ Theorie zu vertreten, die die Inputs und Outputs genau an den Körpergrenzen ansetzt. Tut man dies, so gehören Zustände und Prozesse außerhalb des Körpers definitionsgemäß nicht mehr zum System. Daher können sie auch nicht Zustände oder Prozesse von diesem betrachtet werden.

Wie wir sehen, liegt kein Widerspruch darin, eine funktionalistische Theorie zu vertreten, welche mit dem Paritätsprinzip nicht vereinbar ist. Wenn wir für T eine solche „short-arm“ Theorie einsetzen, folgt die Konklusion:

**Aus der Wahrheit des Funktionalismus folgt nicht die Wahrheit des Paritätsprinzips.**

### 3.2 Folgt der Funktionalismus aus dem Paritätsprinzip?

Die zweite These, die ich in diesem Abschnitt untersuchen möchte, ist folgende:

Wenn das Paritätsprinzip wahr ist, dann ist auch der Funktionalismus wahr.

Zu diesem Thema schreibt Chalmers: „It has sometimes been suggested that the thesis requires functionalism about the mental, where all mental states are defined by the causal roles that they play.“(Chalmers, 2008, XV) Wird also ein Funktionalismus vorausgesetzt, der *alle* mentalen Zustände über ihre kausalen Rollen definiert? Eine solche Variante des Funktionalismus hält Chalmers für unplausibel, da er selber keinen Funktionalismus in Bezug auf phänomenales Bewusstsein vertritt, aber nicht sieht, dass dies den erweiterten Geist gefährdet (vgl. Chalmers, 2008, S. XV). Daraus folgert er, dass nur ein schwacher Funktionalismus benötigt wird.

Ich werde – wie auch Adams und Aizawa (2008, S. 134) – dafür argumentieren, dass das Paritätsprinzip den Funktionalismus überhaupt nicht voraussetzt. Dabei werde ich einem ähnlichen Argumentationsmuster wie schon im letzten Abschnitt folgen:

---

<sup>9</sup>Müller scheint zum Beispiel eine Theorie zu vertreten, welche am ehesten der Möglichkeit 1 entspricht. Möglichkeit 2 würde hingegen am besten zu Rowlands Charakterisierung der Lokalisierungsthese des Internalismus passen.

**Prämisse 1:** Wenn es möglich ist, dass x und y wahr sind und wenn es nicht möglich ist, dass y und z wahr sind, dann folgt aus der Wahrheit von x nicht die Wahrheit von z.

**Prämisse 2:** Es gibt eine Theorie T, für die gilt: Es ist möglich, dass T wahr ist und, dass das Paritätsprinzip wahr ist und es ist nicht möglich, dass T wahr ist und, dass der Funktionalismus wahr ist.

**Konklusion:** Aus der Wahrheit des Paritätsprinzips folgt nicht die Wahrheit des Funktionalismus.

Prämisse 1 sollte aus dem letzten Abschnitt bekannt sein.

**Prämisse 2:** Es gibt eine Theorie T, für die gilt: Es ist möglich, dass T wahr ist und, dass das Paritätsprinzip wahr ist und es ist nicht möglich, dass T wahr ist und, dass der Funktionalismus wahr ist.

Um zu zeigen, dass Prämisse 2 wahr ist, benötige ich ein Beispiel für eine Theorie, die mit dem Paritätsprinzip, jedoch nicht mit dem Funktionalismus vereinbar ist. Sei T zum Beispiel folgende Formulierung der Identitätstheorie:

**Identitätstheorie:** „Jede mentale Eigenschaft bzw. jeder mentale Zustand ist *a posteriori identisch* mit einer physischen Eigenschaft bzw. einem physischen Zustand.“ (Beckermann, 2000, S. 101)

Um zu verstehen, was dies bedeutet und wie dies mit dem Funktionalismus zusammenhängt, werde ich auf das Beispiel des Colaautomaten zurückkommen:

Suchen wir den Zustand, der im Colaautomaten anzeigt, ob schon fünfzig Cent eingeworfen wurden oder nicht, so können wir einen oder mehrere Colaautomaten öffnen. Sind diese Automaten vom elektrischen Typ – wie er oben beschrieben ist – wird man feststellen, dass in allen untersuchten Automaten eine bestimmte Schalterstellung anzeigt, dass schon fünfzig Cent eingeworfen wurden.

Würde man behaupten, dass der Zustand, der anzeigt, ob bereits fünfzig Cent eingeworfen wurden, identisch mit diesem Schalterzustand ist, so wäre man sozusagen Identitätstheoretiker bezüglich der Zustände von Colaautomaten.

Ähnlich sähe es auch in der Hirnforschung aus. Würde man bei allen untersuchten Wesen feststellen, dass bestimmte mentale Zustände immer mit bestimmten physikalischen Zuständen einhergehen, so könnte man behaupten, dass diese identisch wären. Ein mentaler Zustand – zum Beispiel Schmerz – wäre dem zufolge nichts anderes als ein physikalischer Zustand – zum Beispiel das Feuern von C-Fasern. A posteriori wäre eine solche Identitätsaussage, da wir erst in die Welt schauen müssen, ehe wir sie tätigen können. Ist die Identitätstheorie mit dem Paritätsprinzip vereinbar? Ich bin der Meinung, dass dies der Fall ist.

Stellen wir uns zum Beispiel vor, nach dem Öffnen der Colaautomaten würden wir feststellen, dass in den Automaten gar kein Zustand ist, der anzeigt, ob schon fünfzig Cent

eingeworfen wurden oder nicht. Stattdessen ist jeder Automat über das Internet mit einem Server verbunden. Erkennt der Automat, dass fünfzig Cent eingeworfen wurden, so sendet er ein Signal an den Server, welches dies anzeigt. Auf der Festplatte des Servers ist ein Datenbankeintrag, der in einer booleschen Variablen – einer Variablen, die nur zwei verschiedene Werte annehmen kann – speichert, ob in den Automaten schon fünfzig Cent eingeworfen wurden oder nicht. Wird in den Automaten erneut Geld eingeworfen, sendet er eine Anfrage an den Server und bekommt den Wert der entsprechenden Variablen zurück. Je nach Variablenwert und eingeworfener Münze, wird ein bestimmter Schalter für kurze Zeit umgelegt, der die Ausgabe des Automaten bestimmt – siehe Abbildung 4.

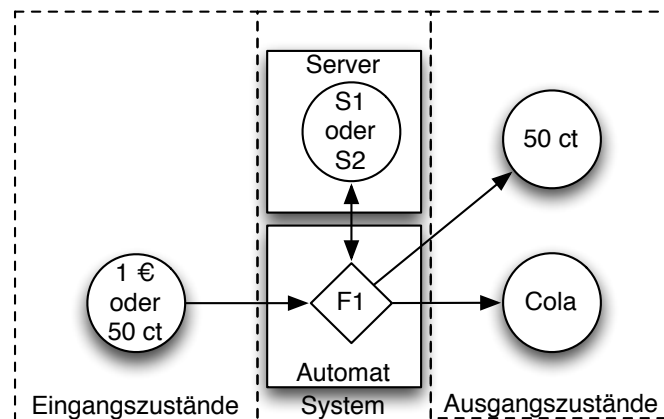


Abbildung 4: Der schematische Aufbau des erweiterten Colaautomatensystems bei Anwendung des Paritätsprinzips.

Ein Identitätstheoretiker bezüglich der Zustände von Colaautomaten müsste nun entscheiden, ob er nur die Vorgänge und Zustände des Colaautomaten betrachtet oder eher das gesamte System, welches Internetverbindung und Server mit einschließt.

Ähnlich könnte es auch dem Identitätstheoretiker mit mentalen Zuständen gehen. Würde man zum Beispiel feststellen, dass die Überzeugung, ein Glas Wein vor sich zu haben, nur dann auftritt, wenn man tatsächlich ein Glas Wein vor sich hat, so müsste der Identitätstheoretiker entscheiden, ob die Überzeugung, ein Glas Wein vor sich zu haben, nur mit einem bestimmten Gehirnzustand identisch ist oder ob das (physische) Glas Wein mit zur betreffenden Überzeugung gehört. Das Paritätsprinzip würde eine Antwort auf diese Frage geben.

**Paritätsprinzip:** Wenn sich der Zustand Z oder der Prozess P teilweise oder komplett außerhalb der körperlichen Grenzen von Subjekt SJ befindet und Z oder P ansonsten im relevanten Sinne kognitiv äquivalent zu internen geistigen Zuständen oder Prozessen von SJ ist, ist Z oder P kognitiv und SJ zuzuschreiben.

In unserem Fall wären die Variablen wie folgt zugeordnet:

**der Zustand M** = der Gesamtzustand aus dem Glas Wein und einem bestimmten neuronalen Zustand des Subjekts SJ

**das Subjekt S** = das Subjekt mit der Überzeugung, ein Glas Wein vor sich zu haben

Wenn wir davon ausgehen, dass der Identitätstheoretiker die physikalischen Zustände, welche mit bestimmten mentalen Zuständen identisch sind, dadurch herausfindet, dass diese physikalischen Zustände *immer* zusammen mit den mentalen Zuständen auftreten, so gehört dem Paritätsprinzip zufolge auch das Glas Wein zur entsprechenden Überzeugung.<sup>10</sup> Wie wir sehen, ist das Paritätsprinzip mit der Identitätstheorie vereinbar. Doch was ist mit Identitätstheorie und Funktionalismus?

Zunächst lässt sich sagen, dass die Identitätstheorie dem Realisiererfunktionalismus ähnlicher ist, als dem Rollenfunktionalismus. In beiden Theorien findet sich die Vorstellung, dass mentale Zustände physikalische Zustände sind. Dennoch gibt es auch hier große Unterschiede.

Es ist unwahrscheinlich, dass man einen physikalischen Zustand findet, der bei Robotern, Außerirdischen, Tieren und Menschen ein und den selben mentalen Zustand darstellt, da diese von ihrer inneren Struktur oder sogar den Materialien, aus denen sie bestehen, völlig unterschiedlich sind. Daher muss die Identitätstheorie entweder bestreiten, dass es mentale Zustände gibt oder sich auf eine bestimmte Gruppe von Wesen beschränken und folglich allen anderen Wesen mentale Zustände absprechen. Ein Realisiererfunktionalist hingegen kann all diesen verschiedenen Wesen mentale Zustände zuschreiben, solange ihre funktionale Struktur hinreichend ähnlich ist. Während die Identitätstheorie nicht behaupten kann, dass Roboter, Außerirdische, Tiere und Menschen mentale Zustände haben, ist dies dem Funktionalisten prinzipiell möglich. Die Identitätstheorie ist also nicht mit dem Funktionalismus vereinbar.

Wir können aus der Prämisse 2 einen wahren Satz bilden, wenn wir für T die obere Formulierung der Identitätstheorie einsetzen. Somit folgt aus Prämisse 1 und 2 die Konklusion:

**Aus der Wahrheit des Paritätsprinzips folgt nicht die Wahrheit des Funktionalismus.**

### 3.3 Sind Funktionalismus und Paritätsprinzip vereinbar?

Ich komme nun zur letzten Frage dieses Kapitels:

Ist es möglich, dass Funktionalismus und Paritätsprinzip zugleich wahr sind?

---

<sup>10</sup>Es sollte offensichtlich sein, dass dies nur ein Gedankenexperiment ist. Faktisch ist natürlich nicht anzunehmen, dass ein Glas Wein immer mit der Überzeugung, ein Glas Wein vor sich zu haben kovariert, da zum Beispiel ein Glas Traubensaft unter entsprechenden Umständen die gleiche Überzeugung hervorbringen kann. Allerdings ist es auch eine Schwäche der Identitätstheorie, dass sie keine multiple Realisierung mentaler Zustände zulässt und daher auch unter den internen Zuständen vermutlich nie einen findet, der in hundert Prozent der Fälle mit einem mentalen Zustand zusammenfällt.

Hierzu schreiben Andy Clark und David Chalmers: „What makes some information count as a belief is the role it plays, and there is no reason why the relevant role can be played only from inside the body.“(Clark und Chalmers, 1998, S. 14)

Mit dem Hinweis darauf, dass die Rolle entscheidend ist, wenn es darum geht, ob eine bestimmte Information eine Überzeugung darstellt, nehmen sie eine klar funktionalistische Position – zumindest Wünsche betreffend – ein. Sie scheinen hier das Paritätsprinzip durch den Funktionalismus begründen zu wollen, was – wie gesehen – nicht unbedingt möglich ist. Dennoch zeigen sie, wie eine Verknüpfung beider Positionen möglich ist. Diese geschieht nämlich einfach durch Ausweitung des betroffenen Systems. Selbst die Kritiker Adams und Aizawa (2008, S. 25) sehen die *Möglichkeit* erweiterter Kognition schon in vielen funktionalistischen Ansätzen enthalten. Auch Weiskopf (2007, S. 266f.) betont ganz richtig, dass das Paritätsprinzip nichts behauptet, was ein Funktionalist unbedingt ablehnen sollte, da dieser sowieso akzeptieren müsse, dass es mitunter sehr ungewöhnliche Realisierungen geistiger Systeme geben könnte.<sup>11</sup>

Multiple Realisierbarkeit ist also eine entscheidende Eigenschaft funktional individuierter mentaler Zustände, die dafür sorgt, dass der Funktionalismus mit dem Paritätsprinzip vereinbar ist. Wir können die Leitfrage dieses Abschnitts wie folgt beantworten:

**Es ist möglich, dass Funktionalismus und Paritätsprinzip wahr sind.**

Aus der Kombination beider Thesen ergibt sich:

### 3.4 Der erweiterte Funktionalismus

Der erweiterte Funktionalismus<sup>12</sup> ist die Kombination der Thesen des Funktionalismus und des Paritätsprinzips.

**Die Konsequenz für das Paritätsprinzip** einer solchen Kombination ist, dass die – ohnehin sehr vage – Bedingung, dass etwas *im relevanten Sinne kognitiv äquivalent* sein muss, durch eine funktionalistische Bedingung ersetzt wird. Eine klare Umformulierung des Paritätsprinzips in diesem Sinne finden wir hier:

„[W]e imagine that exactly the same functional states and processes that are realized in the actual world by those externally located physical elements are now realized by certain internally located physical elements. Having done this, if we then judge that the now internal but previously external processes count as part of a genuinely cognitive system, we are driven to conclude that they did so in the extended case too.“(Wheeler, im Erscheinen, S. 9)

---

<sup>11</sup>So hat zum Beispiel Ned Block in seinem Aufsatz „Troubles with Functionalism“ darauf hingewiesen, dass eine genügend große Anzahl von Menschen, die entsprechend miteinander interagieren würden, die gleiche Funktionalität, wie ein Nervensystem an den Tag legen könnte.(vgl. Beckermann, 2000, S. 165ff.)

<sup>12</sup>Diese Bezeichnung stammt meines Wissens von Wheeler (im Erscheinen).

Was Wheeler hier in Form eines Gedankenexperiments sagen will, ist einfach, dass die funktionale Rolle, die ein externer Zustand für ein kognitives System spielt, für seinen kognitiven Status entscheidend ist.

Betrachten wir zum Beispiel das Colaautomatensystem – inklusive Server –, so realisiert der Wert einer bestimmten Variablen auf dem Server denselben funktionalen Zustand, wie er auch vom Schalterzustand oder von der Hand des Zwerges auf der Schulter des anderen Zwergs realisiert sein könnte. Wir sollten in einem solchen Fall dem erweiterten Colaautomaten-Funktionalismus zufolge nicht darauf achten, wo ein bestimmter funktionaler Zustand realisiert ist.

Ein ähnlicher Fall würde zum Beispiel vorliegen, wenn bestimmte Teile des Gehirns eines Menschen aus dem Körper entfernt und – ohne Einschränkung ihrer Funktion – per Funk neu mit dem Rest des zentralen Nervensystems gekoppelt würden.

Das Paritätsprinzip müsste also wie folgt umformuliert werden:

**funktionales Paritätsprinzip:** Wenn sich der Zustand Z oder der Prozess P teilweise oder komplett außerhalb der körperlichen Grenzen von Subjekt SJ befindet und wenn Z oder P, würden wir Z oder P als Teil des kognitiven Systems S des Subjektes SJ betrachten, in S eine funktionale Rolle einnimmt, die dafür sorgt, dass SJ im mentalen Zustand M ist oder den kognitiven Prozess Q ausführt, gehört Z oder P zum kognitiven System S und ist daher kognitiv und SJ zuzuschreiben.

**Die Konsequenz für den Funktionalismus,** die aus der Kombination mit dem Paritätsprinzip folgt, scheint zu sein, dass kognitive Systeme nicht primär durch ihren Ort, sondern rein funktional bestimmt sind. Dies hat zur Folge, dass wir, Eingabe- und Ausgabezustände betreffend, nicht zu stark auf die Körpergrenze schauen dürfen.

Dies zeigt sich zum Beispiel am erweiterten Colaautomatensystem. In diesem können wir natürlich auch Serveranfragen als Outputs und die zurückgegebenen Werte des Servers als Inputs betrachten – siehe Abbildung 5. Tun wir dies, enthält das betrachtete System keinen Zustand, der anzeigt, ob schon fünfzig Cent eingeworfen wurden oder nicht. Wie aber sollen wir die Grenzen des Systems sonst festlegen?

Im Sinne des funktionalen Paritätsprinzips ließe sich behaupten, dass das System dynamisch so weit gefasst werden sollte, wie es zur Erfüllung einer bestimmten Aufgabe notwendig ist (vgl. Lyre, 2008, S. 14). Es hinge also vom zu erklärenden Mechanismus ab, welche Größe ein System hat.<sup>13</sup>

---

<sup>13</sup>Eine andere Möglichkeit bestünde darin, Kriterien wie folgende aufzustellen (vgl. Clark und Chalmers, 1998, S. 17 und Clark, 2008, S. 79):

Damit ein externer Informationsträger x zum System S dazugehört,

1. muss x für S verlässlich sein und üblicherweise zur Anwendung gebracht werden.
2. sollten Informationen aus x nicht immer von S kritisch überprüft werden.
3. sollten Informationen aus x einfach zugänglich sein, wenn sie gebraucht werden.
4. müssen Informationen in x dort vorhanden sein, da sie schon einmal bewusst von S aufgenommen wurden.

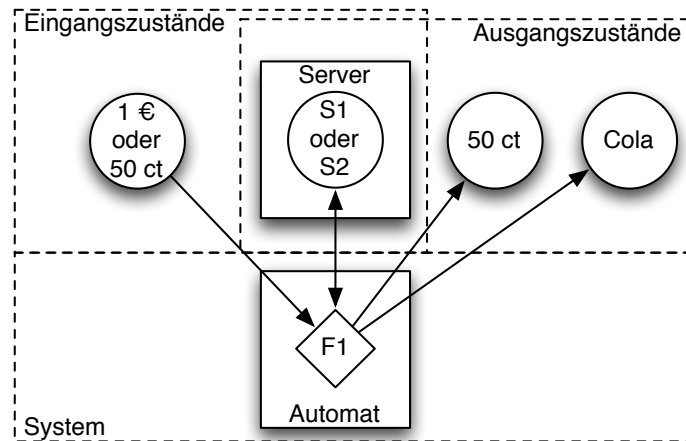


Abbildung 5: Der schematische Aufbau des erweiterten Colaautomatensystems ohne Anwendung des Paritätsprinzips.

Wollen wir wissen, warum das Colaautomatensystem beim Einwurf eines Euros manchmal nur eine Cola und manchmal zusätzlich noch fünfzig Cent ausgibt, betrachten wir das eingeworfene und ausgegebene Geld und die ausgeworfene Cola als Inputs und Outputs. Dadurch gehört der Server offensichtlich mit zum System. Wollen wir hingegen herausbekommen, warum das Colaautomatensystem beim Wert einer Variablen auf dem Server auf die eine oder andere Weise auf den Einwurf von fünfzig Cent reagiert, verkleinern wir das System, sodass nur der Automat als solcher – ohne Internetverbindung und Server – das Gesamtsystem darstellt.<sup>14</sup>

Es ist für den erweiterten Funktionalismus nötig, das betrachtete System nicht mehr durch bestimmte räumliche Grenzen zu bestimmen. Eine Möglichkeit ist, diejenigen Prozesse und Zustände als zum System zugehörig zu betrachten, welche zur Erfüllung einer bestimmten Aufgabe nötig sind. Dies bietet sich an, da wir auf diese Weise nicht das Explanandum – das *zu* Erklärende –, sondern das Explanans – die Erklärung – anpassen müssen. Wir könnten somit noch immer erklären, warum der Einwurf von fünfzig Cent manchmal zu keiner Ausgabe und manchmal zum Auswurf einer Cola führt und müssten

Kriterien *dieser Art* würden höchstwahrscheinlich vom Server erfüllt werden. Eine solche Erweiterung kognitiver Systeme würde dazu führen, dass wir Systeme erhalten würden, die in ihrer Größe relativ festgelegt wären. So würde die Bedingung 1 ausschließen, dass zum Beispiel ein Knoten im Taschentuch in der Funktion eines Erinnerungstriggers (vgl. Rowlands, 2003, S. 174) zum kognitiven System gehören könnte, solange dieses Taschentuch nur einmal für diesen Zweck benutzt werden würde.

<sup>14</sup>An dieser Stelle ergibt sich nun die Frage, ob wir mit einer solchen Systemdefinition überhaupt noch einzelne Subjekte einfangen können. Während zwar in einigen Fällen der Hinweis auf ein kognitives Kernsystem dies zu ermöglichen scheint, sind jedoch auch Systeme vorstellbar, die über einzelne Individuen hinausgehen können.

Wenn zum Beispiel zwei Menschen zusammen einen Aufsatz verfassen, scheint klar, dass die nötigen Prozesse und Zustände zur Erfüllung dieser Aufgabe in *beiden* Menschen und natürlich in einigen Umweltfaktoren – wie zum Beispiel Block, Computer etc. – zu finden sind. Den einen oder anderen Menschen als Kernsystem auszumachen, scheint willkürlich zu sein.

Da mich aber eine ausführliche Beschäftigung mit diesem Thema vom ursprünglichen Ziel dieses Aufsatzes wegführen würde, werde ich hier davon ausgehen, dass ein erweiterter Funktionalismus noch immer eine eindeutige Zuordnung von Zuständen oder Prozessen zu Subjekten zulässt. Anders formuliert: System S lässt sich eindeutig Subjekt SJ zuordnen oder ist sogar mit SJ identisch.

zu den Eingabe- und Ausgabezuständen die Signale des Servers nicht zwangsläufig mit einbeziehen.

Die Loslösung von Systemen, die zunächst räumlich bestimmt sind, führt auch dazu, dass ein erweiterter Funktionalismus ein „long-arm“ Funktionalismus ist (vgl. Wheeler, im Erscheinen, S. 4f.), der als Inputdaten bestimmte Umweltgegebenheiten und als Outputdaten bestimmte Veränderungen in der Umwelt betrachtet. Dies folgt daraus, dass dem „short-arm“ Funktionalismus zufolge Prozesse und Zustände außerhalb des Körpers nicht zum System zählen. Allerdings ist an dieser Stelle Vorsicht geboten. Ein erweiterter Funktionalismus muss nur erlauben, dass einige Informationen außerhalb des Körpers zum System gehören. Man könnte durchaus einen „short-arm“ Funktionalismus bezüglich aller gängigen Sinnesorgane vertreten und dennoch externe Systemerweiterungen zulassen, welche zum Beispiel über Funk oder ähnliche nicht-biologische Mechanismen an den Körper gekoppelt sind.

Als – beinahe triviale – Konsequenz folgt für den Funktionalismus:

**Ein erweiterter Funktionalismus setzt die Möglichkeit voraus, dass die Grenzen des betrachteten Systems S über die Körpergrenzen des entsprechenden Subjekts hinaus gehen.**

Zudem ergeben sich natürlich zwei verschiedene Konsequenzen, je nachdem, ob man einen Rollen- oder einen Realisiererfunktionalismus vertritt.

**Der erweiterte Rollenfunktionalismus** muss behaupten, dass das Gesamtsystem S, welches aus dem Körper des Subjektes SJ und bestimmten externen Komponenten besteht, genau dann mentale Zustände hat, wenn sich das Gesamtsystem in einem bestimmten funktionalen Zustand befindet.

**Der erweiterte Realisiererfunktionalismus** hingegen ist sogar darauf festgelegt, dass wir mentale Zustände, wie Schmerz oder eine bestimmte Überzeugung, auch außerhalb des Körpers finden können. Dies hat die etwas absurd anmutende Konsequenz, dass wir unter den entsprechenden Umständen davon reden können, dass wir einen Wunsch oder einen Teil eines Schmerzes in der Hand halten oder, dass wir eine Überzeugung im Bus verloren haben.<sup>15</sup>

## 4 Ein funktionales Argument für den faktisch erweiterten Geist

Das Paritätsprinzip und damit auch der erweiterte Funktionalismus erlauben lediglich die logische Möglichkeit, dass sich Kognition tatsächlich manchmal über die Grenzen des entsprechenden Subjektes hinaus erstrecken kann. Natürlich wäre es jetzt möglich, dafür

---

<sup>15</sup>Auf ähnliche Konsequenzen weist unter anderem Shapiro (2008, S. 2f.) hin.



zu argumentieren, dass externe Kognition auch mit unseren Naturgesetzen vereinbar ist. Dieser Schritt wird jedoch meist übersprungen, da er nicht mehr nötig ist, wenn erst einmal gezeigt ist, dass es in unserer Welt tatsächlich externe Kognition gibt. Dieser faktische Beweis der Falschheit der internalistischen Lokalisierungsthese wäre auch der Beweis der These des erweiterten Funktionalismus.

**These des erweiterten Geistes:** Es gibt ein Wesen SJ und einen Zustand Z oder einen Prozess P, für die gilt:

Z oder P befindet sich teilweise oder komplett außerhalb der körperlichen Grenzen von Subjekt SJ. Z oder P ist ein kognitiver Zustand oder Prozess des kognitiven Systems S des Wesens SJ oder Teil eines solchen.

#### 4.1 Das Argument der kognitiven Äquivalenz

Es sollen nun verschiedene mögliche Argumentationen für diese These betrachtet werden. Eine davon schreiben Adams und Aizawa (2008, S. 133) einigen Vertretern der These des erweiterten Geistes zu:

„The cognitive equivalence argument has the form of a quasi-syllogism whose major premise strikes us as something like a logical or conceptual truth, namely, that any process that is cognitively equivalent to a cognitive process is itself a cognitive process. The minor premise maintains that this or that process spanning the brain, body, and perhaps environment is cognitively equivalent to a cognitive process.“(Adams und Aizawa, 2008, S. 133)

Aufgeschlüsselt nach Prämissen und Konklusion ergibt sich daraus folgender Quasi-Syllogismus<sup>16</sup>:

**Prämisse 1:** Für alle Prozesse P gilt: Wenn P äquivalent zu einem kognitiven Prozess ist, ist P ein kognitiver Prozess.

**Prämisse 2:** Es gibt einen Prozess P für den gilt: P erstreckt sich über Gehirn, Körper und Umwelt und P ist äquivalent zu einem kognitiven Prozess.

**Konklusion:** Es gibt einen Prozess P für den gilt: P ist ein kognitiver Prozess und P erstreckt sich über Gehirn, Körper und Umwelt.

Im Prinzip ist gegen ein solches Argument nichts einzuwenden. Allerdings soll es in diesem Abschnitt um ein Argument für den erweiterten Geist im Sinne des erweiterten Funktionalismus gehen. Hier jedoch ist nur von allgemeiner und nicht von funktionaler Äquivalenz die Rede.

---

<sup>16</sup>Ein Quasi-Syllogismus zeichnet sich durch die folgende logische Form aus:

**Prämisse 1:** Alle X sind Y.

**Prämisse 2:** a ist X.

**Konklusion:** a ist Y.

## 4.2 Das Argument der funktionalen Äquivalenz

Ein *funktionalistisches* Argument für die These des erweiterten Geistes konstruiert Rupert (2004, S. 41) wie folgt:

**Prämisse 1:** Ein mentaler Zustand des Typs F wird von dem physikalischen Zustand realisiert, der die funktionale Rolle spielt, die charakteristisch (oder metaphysisch individuierend) für F ist.

**Prämisse 2:** Einige Realisierungen von funktionalen Zuständen haben physikalische Komponenten außerhalb des Organismus.

**Prämisse 3:** Ein mentaler Zustand ist ausgedehnt oder enthält alle Komponenten seiner Realisierung.

**Konklusion:** Manche mentalen Zustände erstrecken sich über die Grenzen des Organismus hinaus.

Dieses Argument – besonders Prämisse 1 – basiert auf der Vorstellung des erweiterten Funktionalismus, dass Systeme nicht durch ihren Ort, sondern durch ihre möglichen funktionalen Zustände zu bestimmen sind. Allerdings scheint in Prämisse 3 schon ein Realisiererfunktionalismus vorausgesetzt zu sein. Rollenfunktionalisten können diese Prämisse nicht akzeptieren, da für sie ein mentaler Zustand nur ein Zustand eines gesamten Systems sein kann und somit nicht aus den Komponenten seiner Realisierung besteht.

## 4.3 Ein Argument auf Basis des funktionalen Paritätsprinzips

Eine neutralere Argumentation ist auf Basis des funktionalistischen Paritätsprinzips möglich:

**Prämisse 1 (funktionales Paritätsprinzip):** Wenn sich der Zustand Z oder der Prozess P teilweise oder komplett außerhalb der körperlichen Grenzen von Subjekt SJ befindet und wenn Z oder P, würden wir Z oder P als Teil des kognitiven Systems S des Subjekts SJ betrachten, in S eine funktionale Rolle einnimmt, die dafür sorgt, dass SJ im mentalen Zustand M ist oder den kognitiven Prozess Q ausführt, gehört Z oder P zum kognitiven System S.

**Prämisse 2 (ein Beispiel):** Es gibt ein Subjekt SJ, ein System S und mindestens einen Zustand Z oder einen Prozess P für die gilt:

Der Zustand Z oder der Prozess P befindet sich teilweise oder komplett außerhalb der körperlichen Grenzen des Subjekts SJ. S ist das kognitive System des Subjektes SJ. Z oder P würde, wenn wir Z oder P als Teil des System S betrachten, in S eine funktionale Rolle einnehmen, die dafür sorgt, dass SJ im mentalen Zustand M ist oder einen kognitiven Prozess Q ausführt.

**Konklusion (These des erweiterten Geistes):** Es gibt ein Wesen SJ und einen Zustand Z oder einen Prozess P, für die gilt:

Z oder P befindet sich teilweise oder komplett außerhalb der körperlichen Grenzen von Subjekt SJ. Z oder P ist ein kognitiver Zustand oder Prozess des kognitiven Systems S des Wesens SJ oder Teil eines solchen.

Hat man den erweiterten Funktionalismus erst einmal akzeptiert, ist es nur noch ein kleiner Schritt zur These des erweiterten Geistes. Es wird nur ein Beispiel benötigt, das die Bedingungen der Prämisse 2 erfüllt.

Clark und Chalmers (1998, S. 12f.) führen ein Beispiel an, das von Kritikern und Vertretern der These des erweiterten Geistes immer wieder aufgenommen worden ist. Ich werde nun dieses Beispiel darstellen und mich mit der Frage beschäftigen, ob es sich – dem funktionalen Paritätsprinzip zufolge – um erweiterte Kognition handelt.

#### 4.4 Das Beispiel von Otto und seinem Notizbuch

Erinnern wir uns zurück an Ingas Museumsbesuch:

**Ingas Museumsbesuch:** Inga kommt an einem Plakat vorbei, auf dem steht, dass im Museum eine Kunstausstellung mit Werken Yoko Onos zu sehen ist. Eine Stunde später ist Inga im Museum.

Wir haben gesehen, dass diese Geschichte zwei verschiedene Erklärungen der folgenden Form hat:

**Mentale Erklärung für Ingas Museumsbesuch:** Inga sieht das Plakat. Da sie weiß, welches Datum gerade ist, kommt sie zu dem Schluss, dass die Ausstellung am gleichen Tag stattfindet. Sie wollte schon immer einmal eine Ausstellung von Yoko Ono sehen und hat auch am heutigen Tag nichts anderes vor. Also überlegt sie kurz, wo sich das Museum befindet und geht hin.

**Physikalische Erklärung für Ingas Museumsbesuch:** Sonne strahlt auf das Plakat, das das Licht zurückwirft. Dieses trifft auf Ingas Netzhaut und wird dort in elektrische Signale umgewandelt. Diese Signale erreichen das Gehirn und sorgen dort, zusammen mit anderen Informationen in Form bestimmter elektrochemischer Zustände des Gehirns, dafür, dass am Ende dieses Vorgangs Muskeln in einer bestimmten Art und Weise aktiviert werden. Durch diese Aktivierung bewegt sich Inga in die Richtung des Museums. Während dieser Bewegung sorgt das Gehirn durch ausgleichende Steuerung der Muskeln dafür, dass Inga nicht hinfällt oder vom Weg abkommt.

Mit dem Funktionalismus haben wir auch eine Theorie über den Zusammenhang beider Erklärungen. So wären die elektrochemischen Zustände im Gehirn dem Realisiererfunktionalismus zufolge unter anderem der Wunsch, eine Ausstellung von Yoko Ono zu sehen, die Überzeugung, nichts anderes vor zu haben und eine Überzeugung über die

Richtung, in der das Museum liegt. Der Rollenfunktionalismus würde dem entsprechend diese Zustände als die Realisierer des Wunsches und der Überzeugungen ansehen. Stellen wir uns nun den an Alzheimer erkrankten Otto vor. Dieser schreibt wichtige Informationen schon seit Jahren in sein Notizbuch und liest diese bei Bedarf nach.<sup>17</sup> Nun kommt Otto am selben Plakat wie Inga vorbei. Es kommt zu folgender Situation:

**Ottos Museumsbesuch:** Otto kommt an einem Plakat vorbei, auf dem steht, dass im Museum eine Kunstaussstellung mit Werken Yoko Onos zu sehen ist. Eine Stunde später ist Otto im Museum.

Die physikalische Erklärung dafür sieht wie folgt aus:

**Physikalische Erklärung für Ottos Museumsbesuch:** Sonne strahlt auf das Plakat, das somit Licht zurückwirft. Dieses trifft auf Ottos Netzhaut und wird dort in elektrische Signale umgewandelt. Diese Signale erreichen das Gehirn und sorgen dort, zusammen mit anderen Informationen, in Form bestimmter elektrochemischer Zustände des Gehirns dafür, dass Otto das Notizbuch aus seiner Tasche nimmt, kurz in diesem blättert und dann das Notizbuch wieder einsteckt. Während dieses Blätterns im Notizbuch findet eine komplexe Interaktion zwischen dem Notizbuch und Ottos Gehirn statt. Diese Interaktion besteht aus einer Rückkopplungsschleife, die unter anderem bestimmte Muskelbewegungen Ottos und bestimmte visuelle und taktile Signale einschließt, die vom Notizbuch ausgehen. Jedenfalls führt dieser gesamte Prozess zur Aktivierung bestimmter Muskeln in der Art und Weise, dass sich Otto in die Richtung des Museums bewegt. Während dieser Bewegung sorgt das Gehirn durch ausgleichende Steuerung der Muskeln dafür, dass Otto nicht hinfällt oder vom Weg abkommt.

Nun ergibt sich natürlich die Frage, wie eine mentale Erklärung für Ottos Museumsbesuch aussieht. Zu diesem Thema zitieren Adams und Aizawa (2008) Andy Clark wie folgt:

„Inga’s biological memory systems, working together, govern her behaviours in the functional way distinctive of believing. Otto’s bio-technical matrix (the organism and the notebook) governs his behavior in the same sort of way. So the explanatory apparatus of mental state ascription gets an equal grip in each case, and what looks at first like Otto’s action (looking up the notebook) emerges as part of Otto’s thought.“(Adams und Aizawa, 2008, S. 136<sup>18</sup>)

Clark behauptet also, dass wir Inga und Otto – inklusive Notizbuch – die gleichen mentalen Zustände zurechnen können, auch wenn Otto einige Informationen – zum Beispiel

---

<sup>17</sup>Dieses Beispiel stammt aus Clark und Chalmers (1998, vor allem S. 12f.).

<sup>18</sup>Ein Originalzitat aus Clarks Aufsatz („Active externalism and the extended mind“) ist mir leider nicht möglich, da dieser nicht, wie von Adams und Aizawa angekündigt, im von Robbins und Aydede editierten Sammelband „The Cambridge Handbook of Situated Cognition“ erschienen ist.

den Standort des Museums – nicht in seinem Gehirn gespeichert hat. Clark würde folglich behaupten, dass auf Otto und sein Notizbuch, als System betrachtet, die selbe mentale Erklärung wie auf Inga zutrifft.

**Mentale Erklärung für Ottos Museumsbesuch:** Otto sieht das Plakat. Da er weiß, welches Datum gerade ist, kommt er zu dem Schluss, dass die Ausstellung am gleichen Tag stattfindet. Otto wollte schon immer einmal eine Ausstellung von Yoko Ono sehen und hat auch jetzt nichts anderes vor. Also überlegt er kurz, wo sich das Museum befindet und geht hin.

Der Prozess des Überlegens oder wenigstens seine Realisierung wäre bei Otto unter anderem die komplexe – teils externe – Interaktion mit dem Notizbuch. Die Überzeugung über den Ort des Museums wäre – zumindest teilweise – in Ottos Notizbuch realisiert oder zu finden.

Aber erfüllt Otto zusammen mit seinem Notizblock wirklich die Bedingungen des funktionalen Paritätsprinzips? Zur Überprüfung dieser These sollten wir zunächst die Variablen zuordnen:

**Subjekt SJ** = Otto

**System S** = Ottos kognitives System

**Zustand Z** = der Eintrag in Ottos Notizbuch über den Standort des Museums

**mentaler Zustand M** = eine Überzeugung über den Standort des Museums

Diese Zuordnungen erfüllen die Bedingungen des funktionalen Paritätsprinzips, wenn folgende Sätze wahr sind:

**Bedingung 1:** Der Eintrag in Ottos Notizbuch über den Standort des Museums befindet sich teilweise oder komplett außerhalb der körperlichen Grenzen von Otto.

**Bedingung 2:** Der Eintrag in Ottos Notizbuch über den Standort des Museums würde, wenn wir ihn als Teil von Ottos kognitivem System betrachten, in diesem eine funktionale Rolle einnehmen, die dafür sorgt, dass Otto eine Überzeugung über den Standort des Museums hat.

Die erste Bedingung ist erfüllt und bedarf keiner weiteren Betrachtung. Bedingung 2 hingegen ist problematischer. Eine Überprüfung dieses Satzes setzt zunächst eine Theorie darüber voraus, welche funktionale Rolle realisiert sein muss, damit ein System eine bestimmte Überzeugung hat.

Clark und Chalmers (1998, S. 13) sehen eine klare Analogie zwischen Inga und Otto, da in beiden Fällen die entsprechende Information zuverlässig dem Bewusstsein zur Handlungsplanung zur Verfügung steht. Dies scheint alleine nicht als funktionale Definition auszureichen, da ansonsten der Inhalt jedes Buches, welches jemand ständig bei sich hat, auch Inhalt des Gedächtnisses wäre.

Es müsste also im Sinne von Clark und Chalmers (1998, S. 17) und Clark (2008, S. 79), hinzugefügt werden, dass die Informationen in Ottos Notizbuch auch deshalb zu seinem Gedächtnis zählen, da Otto diese Informationen schon einmal bewusst aufgenommen hat. Nun stößt man allerdings auf das Problem, dass Informationen, die zum Beispiel durch einen verrückten Neurowissenschaftler in das Gehirn einer Person „eingepflanzt“ worden wären, ohne dass diese Person diese Informationen bewusst aufgenommen hätte, keine Überzeugungen sein könnten. Dies scheint kontraintuitiv zu sein. Müssen wir also anerkennen, dass ein evangelikaler Christ<sup>19</sup>, der seine Bibel immer bei sich trägt, den Inhalt der gesamten Bibel zu seinen Überzeugungen zählen kann?

Diese und ähnliche Probleme tauchen auf, wenn man versucht herauszubekommen, ob ein physikalischer Zustand eine bestimmte funktionale Rolle einnimmt. Ich werde an dieser Stelle nicht weiter nach einer korrekten funktionalen Definition für Überzeugungen suchen, da dies den Umfang dieses Aufsatzes sprengen würde.<sup>20</sup>

#### **4.5 Das Problem des Fehlens einer funktionalistischen Theorie**

Wir können zwar in gewisser Weise verstehen, was Clark und Chalmers (1998, S. 13f.) damit meinen, wenn sie sagen, dass Ottos Notizbuch die selbe Rolle einnimmt, wie die Überzeugung oder ihre Realisierung in Ingas Gehirn. Ob diese funktionale Ähnlichkeit wirklich hinreichend ist, bleibt hingegen fraglich.

Versucht man auf Basis des funktionalen Paritätsprinzips dafür zu argumentieren, dass es in der Welt einen faktisch erweiterten Geist gibt, so setzt dies nicht nur ein Beispiel für einen Zustand oder Prozess der in Frage kommt voraus. Mindestens ebenso wichtig ist eine Theorie, die zeigen kann, dass dieser Zustand oder Prozess im relevanten Sinne funktional äquivalent ist.

**Ein Argument für den erweiterten Geist auf Basis des funktionalen Paritätsprinzips scheitert in Folge des Fehlens einer funktionalen Theorie, die es erlaubt, mögliche Beispiele auf ihre funktionale Gleichheit zu überprüfen.**

### **5 Ein Argument gegen (Realisierer von) Überzeugungen in Ottos Notizbuch**

Im Argument für den faktisch erweiterten Geist auf Basis des funktionalen Paritätsprinzips konnte bisher *nur* nicht gezeigt werden, dass Prämisse 2 tatsächlich wahr ist. Das Beispiel von Otto scheitert daran, dass keine funktionalistische Theorie zur Verfügung steht, die die funktionale Äquivalenz des Notizbuchs zum Gehirn aufzeigen kann.

<sup>19</sup>Evangelikale Christen glauben an die Irrtumsfreiheit der Bibel.

<sup>20</sup>Dies ergibt sich unter anderem daraus, dass auch die mentalen Zustände, die in einer solchen Definition vorkommen würden, selbst eine Definition benötigen. Und vermutlich wären auch diese Definitionen nicht frei von Referenzen zu anderen – noch nicht definierten – mentalen Zuständen. Ein verwandtes Problem findet sich in Levin (2009, im Abschnitt „Functionalism and Holism“).

O'Brien (1998), Rupert (2004), Weiskopf (2007), Shapiro (2008), Adams und Aizawa (2008) wollen jedoch mehr zeigen. Ihnen geht es um den Beweis, dass das Einbeziehen des Notizbuchs in Ottos kognitives System nicht rechtfertigt, Otto neue Überzeugungen zuzuschreiben. Diesen wollen sie liefern, indem sie bestreiten, dass ein Eintrag in Ottos Notizbuch funktional äquivalent zu bestimmten Gehirnzuständen Ingas sein kann.

## 5.1 Der Aufbau der Argumentation

Ihre Argumentation lässt sich auf zwei verschiedene Arten und Weisen darstellen:

**Auf Basis des Rollenfunktionalismus** ist es ein Argument gegen Realisierer von Überzeugungen in Ottos Notizbuch.

**Auf Basis des Realisiererfunktionalismus** handelt es sich um ein Argument gegen Überzeugungen in Ottos Notizbuch.

### Argument gegen Realisierer von Überzeugungen in Ottos Notizbuch

**Prämisse 1:** Die Menge aller Zustände mit der funktionalen Eigenschaft E in Ottos kognitivem System ohne Notizbuch ist identisch mit der Menge aller Zustände mit der funktionalen Eigenschaft E in Ottos kognitivem System mit Notizbuch.

**Prämisse 2:** Für alle beobachteten Zustände M gilt: Wenn M eine Überzeugung realisiert, hat M die funktionale Eigenschaft E.

**Prämisse 3:** Für alle Zustände Z und alle beobachteten Zustände M gilt: Wenn jedes M, welches eine Überzeugung realisiert, eine funktionale Eigenschaft hat, dann realisiert Z nur eine Überzeugung, wenn es diese funktionale Eigenschaft hat.<sup>21</sup>

**Zwischenschritt** (aus Prämisse 2 und 3): Für alle Zustände Z gilt: Z realisiert nur eine Überzeugung, wenn Z die funktionale Eigenschaft E hat.

**Konklusion** (aus Prämisse 1 und dem Zwischenschritt): Die Menge aller Überzeugungen von Ottos kognitivem System ohne Notizbuch ist identisch mit der Menge aller Überzeugungen von Ottos kognitivem System mit Notizbuch.

### Argument gegen Überzeugungen in Ottos Notizbuch

**Prämisse 1:** Die Menge aller Zustände mit der funktionalen Eigenschaft E in Ottos kognitivem System ohne Notizbuch ist identisch mit der Menge aller Zustände mit der funktionalen Eigenschaft E in Ottos kognitivem System mit Notizbuch.

**Prämisse 2:** Für alle beobachteten Zustände M gilt: Wenn M eine Überzeugung ist, hat M die funktionale Eigenschaft E.

---

<sup>21</sup>Prämisse 3 wird zwar meist nicht explizit vertreten. Dennoch muss sie oder zumindest eine gleichwertige Prämisse vertreten werden, um zur Konklusion zu gelangen. Dies gilt natürlich auch für Prämisse 3 des Arguments gegen Überzeugungen in Ottos Notizbuch.

**Prämisse 3:** Für alle Zustände Z und alle beobachteten Zustände M gilt: Wenn jedes M, welches eine Überzeugung ist, eine funktionale Eigenschaft hat, dann ist Z nur eine Überzeugung, wenn es diese funktionale Eigenschaft hat.

**Zwischenschritt** (aus Prämisse 2 und 3): Für alle Zustände Z gilt: Z ist nur eine Überzeugung, wenn Z die funktionale Eigenschaft E hat.

**Konklusion** (aus Prämisse 1 und dem Zwischenschritt): Die Menge aller Überzeugungen in Ottos kognitivem System ohne Notizbuch ist identisch mit der Menge aller Überzeugungen in Ottos kognitivem System mit Notizbuch.

Wie wir sehen, ist eine solche Argumentation sehr beschränkt, da sie natürlich nicht zeigen kann, dass die These des erweiterten Geistes falsch ist. Sie kann nicht einmal zeigen, dass Ottos Notizbuch *in keiner Weise* dazu beiträgt, Ottos kognitives System zu erweitern. Die einzige Schlussfolgerung die hier gezogen werden kann ist, dass Ottos Notizbuch (und natürlich auch Notizbücher anderer Personen in ähnlichen Fällen), im Gegensatz zu Teilen seines Gehirns, kein Speicher oder Teilspeicher von Erinnerungen sein kann.

Trotz der begrenzten Reichweite dieses Arguments werde ich es dennoch ausführlicher besprechen, da der Fall von Otto und seinem Notizbuch ein zentrales Beispiel in der Diskussion über den erweiterten Geist darstellt. Zudem macht es Prämisse 3 besonders leicht, Argumente gegen weitere Beispiele dieser Art zu konstruieren.

## 5.2 Mögliche Wahrmacher der Prämissen 1 und 2

Welche funktionalen Eigenschaften E machen die Prämissen 1 und 2 wahr? Oder anders gefragt: Welche funktionale Eigenschaft lässt sich bei allen bisher untersuchten Überzeugungen oder ihren Realisierern, jedoch nicht bei Otto und seinem Notizbuch feststellen? Es sind viele Eigenschaften vorgeschlagen worden, von denen ich nun einige darstellen werde.

**Kausale Aktivität und funktionale Integration** sind für bisher beobachtete Überzeugungen zentral. So sind in biologischen Gedächtnissen Informationen völlig anders gespeichert, als in einem Notizbuch. Während bisher untersuchte kognitive Systeme Informationen in komplexen Netzwerken aus Neuronen speichern, sind Informationen in einem Notizbuch in Form von – meist sprachlichen – Zeichen vorhanden.

Neuronale Netzwerke speichern Informationen distributiv. Das bedeutet, dass Überzeugungen (oder ihre Realisierer) in Gehirnen oder künstlichen neuronalen Netzwerken auf die Gesamtheit der Synapsen verteilt sind (vgl. Lyre, 2002, S. 144). Somit können diese nicht von anderen Überzeugungen getrennt betrachtet werden. Zudem werden Informationen in solchen Netzwerken parallel – gleichzeitig – und nicht sequenziell – nacheinander – verarbeitet.

„[A]ll the information that is encoded in this fashion is causally active *whenever* that



network responds to an input.“(O’Brien, 1998, S. 3) Eine solche Informationsspeicherung in Gehirnen hat also zur Folge, dass Überzeugungen kausal aktiv auf andere Überzeugungen, Handlungen oder Wünsche einwirken. So schreibt Weiskopf (2007):

„Beliefs are, as I will say, normally informationally integrated with, and updated in concert with, other beliefs (and further mental states of the subject, such as desires).“ (Weiskopf, 2007, S. 268)

Zur Veranschaulichung dieses Phänomens soll folgendes Beispiel dienen.

Nehmen wir an, Otto war vor drei Jahren das erste mal im Museum der Stadt. Damals hat er sich die Adresse in sein Notizbuch geschrieben. Doch vor einem Jahr las er in der Zeitung, dass das alte Museumsgebäude renoviert wird und dass daher das Museum für zwei Jahre in ein anderes Gebäude verlegt werden soll. Otto notierte sich auch die neue Adresse, allerdings ohne die alte Information zu ändern. Will Otto nun ins Museum, hängt es davon ab, ob er sein Notizbuch von vorne nach hinten oder von hinten nach vorne durchsucht, zu welcher Adresse er geht. Inga hingegen hätte, wenn sie die selben Inputs – durch den ersten Museumsbesuch vor drei Jahren und das Lesen der Zeitung zwei Jahre später – bekommen hätte, ihre alte Überzeugung automatisch angepasst. Es ist natürlich denkbar, dass Otto ein System hat, welches ihm erlaubt, solche Fehler zu vermeiden (vgl. Weiskopf, 2007, S. 270f.). Es ist jedoch sehr unwahrscheinlich, dass er tatsächlich ein System benutzt, das ihm erlaubt, Inkonsistenzen in seinen Überzeugungen und Wünschen auf die Weise zu vermeiden, wie es in neuronalen Netzen der Fall ist.

Es lässt sich sagen, dass die funktionale Eigenschaft von neuen Überzeugungen automatisch aktualisiert zu werden, eingesetzt ist Prämisse 1 und 2, diese Prämissen wahr macht.

**Der Generierungs-Effekt** ist eine weitere Eigenschaft neuronaler Gedächtnisse, auf die Rupert (2004, S. 33f.) und später auch Adams und Aizawa (2008, S. 139) hingewiesen haben.

So erinnern sich Menschen signifikant häufiger an vorgegebene Wörter, wenn sie beim Lernen dieser Wörter etwas aus diesen generieren mussten. Ist eine Versuchsperson angewiesen einen Satz oder ein anderes Wort (Gegenteil, Synonym, Assoziation, Reim) zu einem vorgegebenen Wort zu sagen, ist die Wahrscheinlichkeit höher, dass sie sich hinterher an das Vorgegebene erinnert, als wenn schon solchen ein Satz, ein Gegenteil, ein Synonym oder ein Reim vorgegeben worden wäre.

Es ist jedoch sehr unwahrscheinlich, dass bei Otto und seinem Notizbuch ein Generierungseffekt vorzufinden ist. Otto hat die Möglichkeit, alle Wörter aufzuschreiben, unabhängig davon, ob er sie weiterverarbeitet oder nicht.

Die hier untersuchten Überzeugungen – zum Beispiel die Überzeugung, dass einem das Wort „Pferd“ gezeigt wurde – haben die funktionale Eigenschaft, mit geringerer Wahrscheinlichkeit vergessen zu werden, wenn beim Erlangen der Überzeugung diese zur Ge-

nerierung eines Outputs genutzt wurde (vgl. Engelkamp und Zimmer, 1994, 69f.). Diese Eigenschaft hätten in Ottos Notizbuch gespeicherte oder realisierte Überzeugungen nicht.

**Das Vergessen** als solches ist im erweiterten Ottosystem anders als in „normalen“ Gedächtnissen. So weist Shapiro (2008, S. 21) darauf hin, dass neuronale Langzeitgedächtnisse mit der Zeit vergessen. Diesem Vergessen kann aber durch regelmäßigen Abruf der gespeicherten Informationen entgegengewirkt werden (vgl. Turkington und Harris, 2001, S. 92f.).

In Ottos Notizbuch ist ein solcher Effekt jedoch nicht vorhanden. Auch wenn die Seiten vielleicht mit der Zeit verblassen, kann Otto diesem Effekt nicht alleine dadurch entgegenwirken, dass er die Einträge von Zeit zu Zeit liest.

Bisher beobachtete Überzeugungen haben also die funktionale Eigenschaft, mit der Zeit vergessen, aber durch regelmäßiges Nutzen dieser Überzeugungen behalten zu werden. Informationen in Ottos Notizbuch haben diese Eigenschaft nicht.

**Der Primacy-Recency-Effekt** ist das letzte Beispiel dieser Art, auf das ich hier eingehen will. Es stammt von Adams und Aizawa (2008, S. 137).

Versuche, bei denen es, ähnlich wie bei den Versuchen zum Generierungseffekt, darum geht, sich Wörter zu merken, haben Folgendes gezeigt. Versuchspersonen ohne Hilfsmittel können sich die zuerst und die zuletzt genannten Wörter am besten merken (vgl. Turkington und Harris, 2001, S. 188 und S. 197). Werden einer Versuchsperson zwanzig Wörter jeweils im Abstand von zwei Sekunden vorgelesen, so ist statistisch gesehen zu erwarten, dass sie bei einer anschließenden Frage, welche Wörter genannt wurden, sich am besten an die Wörter am Anfang und am Ende erinnern. Dies ist der Primacy-Recency-Effekt.

Otto hingegen könnte, wenn er genügend Übung im Umgang mit seinem Notizbuch hat, sich alle genannten Wörter aufschreiben.

Bisher untersuchte Überzeugungen haben die funktionale Eigenschaft, mit geringerer Wahrscheinlichkeit vergessen zu werden, wenn sie am Anfang oder Ende einer Reihe zu merkender Informationen auftauchen.

Ich will an dieser Stelle nicht beurteilen, ob tatsächlich *alle* bisher untersuchten Gedächtnisse die genannten Eigenschaften haben. Daher sei den Kritikern des Ottobeispiels zugestanden, dass die hier genannten funktionalen Eigenschaften tatsächlich die Prämissen 1 und 2 wahr machen können, wenn sie für E eingesetzt werden.

Doch alleine aus diesen Beispielen lässt sich nichts folgern. Zwar können sie zeigen, dass nicht-erweiterte Gedächtnisse funktional anders sind als durch Notizbücher erweiterte. Wenn dies allerdings vor dem Hintergrund der Theorie des erweiterten Funktionalismus zu dem Schluss führen soll, dass Notizbücher tatsächlich nicht geeignet sind, um Gedächtnisse zu erweitern, wird mindestens eine weitere Annahme benötigt. In der oben angeführten Argumentation ist dies Prämisse 3.

### 5.3 Die Prämisse 3 und der Psychofunktionalismus

**Prämisse 3:** Für alle Zustände Z und alle beobachteten Zustände M gilt: Wenn jedes M, welches eine Überzeugung ist oder realisiert, eine funktionale Eigenschaft hat, dann ist oder realisiert Z nur eine Überzeugung, wenn es diese funktionale Eigenschaft hat.

Wurde zum Beispiel bei jeder bisher beobachteten (Realisierung einer) Überzeugung festgestellt, dass diese zumindest prinzipiell vergessen werden kann, so gehört – Prämisse 3 zufolge – die funktionale Eigenschaft vergessen werden zu können zur Definition einer Überzeugung dazu.

Eine solche Behauptung erinnert stark an eine Version des Funktionalismus, die sich „Psychofunktionalismus“ oder „empirischer Funktionalismus“ nennt. Diese These lässt sich wie folgt beschreiben:

„What is distinctive about psycho-functionalism is its claim that mental states and processes are just those entities, with just those properties, postulated by the best scientific explanation of human behaviour.“ (Levin, 2009)

Der Psychofunktionalismus behauptet also, dass die besten wissenschaftlichen Theorien über das *menschliche* Verhalten festlegen, welche funktionalen Eigenschaften Zustände oder Prozesse haben müssen, damit es sich bei ihnen um bestimmte kognitive Zustände oder Prozesse oder deren Realisierungen handelt. Damit unterscheidet sich der Psychofunktionalismus in folgenden Punkten von Prämisse 3:

- Prämisse 3 betrifft nur Überzeugungen. Der Psychofunktionalismus hingegen beansprucht, alle kognitiven Zustände und Prozesse definieren zu können. Dies hat zur Folge, dass es für den Psychofunktionalismus möglich ist, Überzeugungen gar nicht mehr zu den mentalen Zuständen zu zählen.
- In Prämisse 3 ist nicht vorausgesetzt, dass sich die Beobachtungen in wissenschaftlichen Theorien wiederfinden müssen.
- Prämisse 3 macht keine explizite Referenz auf menschliche Kognition und ist damit in diesem Punkt liberaler als der Psychofunktionalismus.

In Folge dieser Unterschiede schließt der Psychofunktionalismus Prämisse 3 nicht ein und auch Prämisse 3 setzt den den Psychofunktionalismus nicht voraus. Dennoch sind beide Thesen von folgender Analogie betroffen:

Wie ich schon festgestellt habe, ist auch der Begriff „Pumpe“ funktional definiert. Ob ein Gegenstand eine Pumpe ist, hängt davon ab, ob er Flüssigkeiten oder Gasen potentielle Energie in Form von Druck oder kinetische Energie zuführen kann. Eine solche Definition kann, wie wir gesehen haben, sehr unterschiedliche Gegenstände wie Luftpumpen, Kolbenpumpen und sogar das Herz umfassen. Es sind also keine Detailkenntnisse der Funktionsweise notwendig, um zu erkennen, ob etwas eine Pumpe ist.

Es wäre möglich, so etwas wie einen „empirischen Funktionalismus für Pumpen“ zu vertreten. Die zentrale These eines solchen wäre: Etwas ist eine Pumpe, wenn es die Funktion erfüllt, die unsere besten wissenschaftlichen Theorien über die Funktionsweise von Pumpen postulieren. Stellen wir uns nun eine Welt vor, in der bisher nur Pumpen entdeckt oder erfunden wurden, die Flüssigkeiten transportieren oder unter Druck setzen. Es gäbe also keine Blasebälge, Luftpumpen und andere Pumpen, deren Funktion es ist, einem Gas potentielle oder kinetische Energie zuzuführen. In einer solchen Welt müsste der empirische Pumpenfunktionalist behaupten, dass etwas nur eine Pumpe ist, wenn es Flüssigkeiten die entsprechende Energie zuführt.

Damit kämen wir zu den Einwänden gegen den Psychofunktionalismus und Prämisse 3.

#### 5.4 Ein Einwand gegen Prämisse 3 und den Psychofunktionalismus

Nehmen wir an, in der gerade skizzierten Welt gibt es zwar schon lange die Idee, etwas zu bauen, was in der Lage ist, Gase zu befördern. Diese ist jedoch aufgrund einiger technischer Schwierigkeiten nie umgesetzt worden. Erzählt nun ein Erfinder einem empirischen Pumpenfunktionalisten von dieser Idee, so könnte dieser zwar zugestehen, dass es sich um eine interessante Idee handelt. Erzählt der Erfinder jedoch, dass er seine Erfindung „Luftpumpe“ nennen will, so würde der empirische Pumpenfunktionalist behaupten, dass es sich beim vorgeschlagenen Gerät nicht um eine Pumpe handeln kann, da Pumpen nur mit Flüssigkeiten funktionieren.

Wir sollten aus dieser Analogie nicht zu viel ableiten, da sich unsere Intuitionen über Pumpen natürlich von denen in der beschriebenen Welt unterscheiden können. Dennoch scheint es, als würde der empirische Pumpenfunktionalist einen Fehler machen. Dieser zeigt sich darin, dass in seiner These („Etwas ist eine Pumpe, wenn es die Funktion erfüllt, die unsere besten wissenschaftlichen Theorien über die Funktionsweise von Pumpen postulieren.“) schon der Begriff „Pumpe“ enthalten ist. Er muss folglich die Möglichkeit voraussetzen, zu entscheiden ob etwas eine Pumpe ist oder nicht, ohne dabei schon auf wissenschaftliche Theorien zurück zu greifen.

Ähnlich geht es dem Psychofunktionalisten oder dem Vertreter der Prämisse 3: Natürlich sind Außerirdische aus Silicium oder Roboter mit pneumatischen Gehirnen nicht wissenschaftlich untersucht oder auch nur beobachtet worden. Dennoch können wir uns im Gedankenexperiment solche Wesen vorstellen und auch entscheiden, ob diese geistige Zustände haben würden. Die Frage ist nun, ob sich bei solchen Wesen tatsächlich die oben genannten Beispiele (wie der Generierungs- oder ein Primacy-Recency-Effekt) feststellen lassen müssen, damit sie Überzeugungen haben. Die These, dass alle diese Effekte tatsächlich notwendig sind, damit ein Wesen Überzeugungen hat, lehnen Wheeler (im Erscheinen, S. 11 und S. 17) und Clark (2008, S. 114f.) ab, da solche Bedingungen kontraintuitive Ergebnisse liefern.

Der Psychofunktionalismus und die Prämisse 3 werden unseren Intuitionen, dass kognitive Wesen im Detail auch anders funktionieren können, nicht gerecht. Das Argument

gegen (Realisierungen von) Überzeugungen in Ottos Notizbuch scheitert also in dieser Form, da eine entscheidende Prämisse unplausibel ist. Dennoch muss diese natürlich durch eine andere Annahme darüber, wann etwas eine Überzeugung ist, ersetzt werden.

### 5.5 Der Common-Sense Funktionalismus die neue Prämisse 3

Wenn der empirische Funktionalismus unsere Fähigkeit voraussetzt Überzeugungen zu erkennen, scheint es naheliegend, die Faktoren anhand derer wir dies erkennen können, explizit zu machen. Oder anders gesagt:

„Most of us understand [mental state terms] and so any account of their meaning should respect this fact.“(Braddon-Mitchell und Jackson, 1996, S. 79)

Genau dies tut der Common-Sense Funktionalismus. Wann ein System in einem mentalen Zustand ist, hängt dieser Theorie zufolge nicht von wissenschaftlichen Ergebnissen, sondern von unseren Alltagsintuitionen ab. Es ist also nicht nötig, Kognition- oder Neurowissenschaftler zu sein, um feststellen zu können, ob eine Person eine Überzeugung oder einen Wunsch hat. Ein intuitives Verständnis und ein wenig Erfahrung im Umgang mit kognitiven Wesen reichen aus.

Wenn wir einen Ersatz für Prämisse 3 benötigen, so würde ich – im Sinne des Common-Sense Funktionalismus – folgende vorschlagen:

**Neue Prämisse 3:** Für alle Zustände  $Z$  und alle vorstellbaren Zustände  $M$  gilt: Wenn jedes  $M$ , welches eine Überzeugung ist oder realisiert, eine funktionale Eigenschaft hat, dann ist oder realisiert  $Z$  nur eine Überzeugung, wenn es diese funktionale Eigenschaft hat.

Damit eine funktionale Eigenschaft zur Definition einer Überzeugung dazugehört, darf es nicht denkbar sein, dass ein Zustand die funktionale Eigenschaft  $E$  nicht hat und trotzdem eine Überzeugung realisiert oder ist.

Es ist natürlich nicht zu erwarten, dass wir aus einer solchen Analyse überraschende Ergebnisse erhalten. Ein Generierungseffekt wird vermutlich nicht notwendig für Überzeugungen sein. Dies hat natürlich eine deutlich liberalere Praxis im Zuschreiben von Überzeugungen zur Folge. Einer solchen folgt auch Clark, wenn er schreibt:

„[T]aken a single, integrated system, Otto-and-the-notebook exhibit enough of the central features and dynamics of a normal agent having (amongst others) the dispositional belief that MOMA is on 53rd St. to warrant treating him as such.“(Clark, 2005, S. 7)

Ob in unserem Beispiel Otto tatsächlich eine Überzeugung über den richtigen Standort des Museums hat, kann ich – wie schon oben gesagt – nicht entscheiden. Auch die Diskussion, ob die oben genannten Effekte tatsächlich nicht notwendig sind, würde an

dieser Stelle den Rahmen des Aufsatzes sprengen. Es lässt sich aber sagen, dass ein erweiterter Common-Sense Funktionalismus deutlich wahrscheinlicher zur These des erweiterten Geistes führt, als ein erweiterter Psychofunktionalismus.<sup>22</sup>

Hat der Common-Sense Funktionalismus das Ziel, unsere Alltagsüberzeugungen über mentale Zustände in eine Theorie zu fassen, so scheint es unplausibel zu sein, dass es sich bei ihm um einen „short-arm“ Funktionalismus handelt. Wenn wir im Alltag über mentale Zustände reden, so tun wir dies nämlich nicht, indem wir Behauptungen darüber aufstellen, welche Signale im Körper oder gar im zentralen Nervensystem ankommen. Viel mehr scheinen die Inputfaktoren Umweltbedingungen und die Outputfaktoren beobachtbaren Veränderungen am Wesen oder in seiner Umwelt zu sein.

Es ist nicht entscheidend, dass Licht einer bestimmten Wellenlänge die Retina erreicht, sondern, dass vor dem Subjekt ein Schrank steht. Es ist nicht entscheidend welche Signale das Hirn an die Muskeln schickt, sondern, dass das das Subjekt zwinkert, läuft oder etwas wirft.

Ein Common-Sense Funktionalismus ist also zumindest kein strenger „short-arm“ Funktionalismus, sondern höchstens eine Mischform, die die Input- und Outputgrenzen nicht notwendigerweise an der Körpergrenze ansetzt. Ein Common-Sense Funktionalismus muss kein erweiterter Funktionalismus sein. Eine Erweiterung scheint aber höchstens ein kleiner Schritt zu sein.

Noch ein paar Worte zum Ende dieses Abschnitts: Ich will keineswegs behaupten, dass der Generation- oder der Primacy-Recency-Effekt nicht untersucht werden sollten. Eine so genannte microfunktionale Untersuchung von Tieren und Menschen kann durchaus interessant sein und wertvolle Erkenntnisse bringen. Ich behaupte nur, dass geistige Phänomene wie Überzeugungen und Wünsche nicht durch empirische Forschung *definiert* werden sollten. Eine solche Ansicht setzt, wie ich gezeigt habe, nicht schon die These des erweiterten Geistes voraus. Damit komme ich zu einem anderen Ergebnis als zum Beispiel Adams und Aizawa (2008, S. 140).

## 6 Eine Rekapitulation

Ich habe nun das Verhältnis vom Funktionalismus zum Paritätsprinzip untersucht und mich auch mit der Frage beschäftigt, ob Otto – inklusive Notizbuch – als Anwendungsfall der Kombination beider Positionen zeigt, dass die These des erweiterten Geistes wahr ist. Eine vollständige Wiedergabe der Debatte hätte allerdings den Umfang dieser Arbeit gesprengt, sodass viele Aspekte dieses Themengebiets nicht untersucht werden konnten.<sup>23</sup> Dennoch bin ich zu einigen Ergebnissen gekommen, die ich an dieser Stelle

---

<sup>22</sup>Zugegebenermaßen ist dies nicht notwendigerweise so, da sich der Psychofunktionalismus nur darauf festlegt, dass die *aktuell* beste Theorie entscheidend ist. Es legt natürlich nichts fest, dass eine solche Theorie immer die internalistische Lokalisierungsthese enthalten muss. Haben zum Beispiel Clark und Chalmers recht, so liefert eine Herangehensweise im Sinne des erweiterten Geistes tiefere, einheitlichere und nützlichere Erklärungen (vgl. Clark und Chalmers, 1998, S. 14).

<sup>23</sup>So ist im Aufsatz von Clark und Chalmers nicht nur das Beispiel Ottos, sondern auch eines angeführt worden, in dem es um gedankliche und praktische Drehung von Steinen im Videospiel Tetris geht (vgl.

kurz zusammenfassen will:

**Der Funktionalismus** lässt sich wie folgt darstellen:

**minimaler Funktionalismus:** Für alle mentalen Zustände  $M$  und alle Systeme  $S$  gilt:  $S$  ist genau dann im mentalen Zustand  $M$ , wenn es sich im funktionalen Zustand  $Z$  befindet.

Innerhalb des Funktionalismus gibt es folgende Möglichkeiten:

**Rollenfunktionalismus:** Minimaler Funktionalismus und  $M = Z$

**Realisiererfunktionalismus:** Minimaler Funktionalismus und  $M$  ist das, was im System  $S$  die Rolle  $Z$  erfüllt.

**Das Paritätsprinzip** ist die Behauptung, dass es logisch möglich ist, dass Kognition über die Grenzen des Körpers hinausgeht:

**Paritätsprinzip:** Wenn sich der Zustand  $Z$  oder der Prozess  $P$  teilweise oder komplett außerhalb der körperlichen Grenzen von Subjekt  $SJ$  befindet und  $Z$  oder  $P$  ansonsten im relevanten Sinne kognitiv äquivalent zu internen geistigen Zuständen oder Prozessen von  $SJ$  ist, ist  $Z$  oder  $P$  kognitiv und  $SJ$  zuzuschreiben.

**Das Verhältnis von Funktionalismus und Paritätsprinzip:**

**Der Funktionalismus setzt nicht das Paritätsprinzip voraus**, da die Möglichkeit, die Grenzen des Systems genau an der Körpergrenze anzusetzen, nicht ausgeschlossen ist.

**Das Paritätsprinzip setzt den Funktionalismus nicht voraus**, da es auch mit anderen Theorien über die Natur geistiger Zustände (z.B. Identitätstheorie) vereinbar ist.

**Es ist möglich, dass Funktionalismus und Paritätsprinzip wahr sind**, da der Funktionalismus die Grenze des betrachteten Systems betreffend neutral ist, sodass diese vom Paritätsprinzip festgelegt werden kann.

---

Clark und Chalmers, 1998, S. 7f.). Natürlich sind auch hier Einwände erhoben worden, zum Beispiel von Adams und Aizawa (2008, S. 141ff.).

Auch die Frage, ob sich Ottos funktionale Organisation nicht auch im Sinne eines erweiterten Common-Sense Funktionalismus von der Ingas unterscheidet, wurde nicht behandelt. Zu diesem Thema ist zum Beispiel der Einwand möglich, dass Otto vor dem Zugriff auf sein Notizbuch, die Überzeugung, die gewünschte Information in diesem Notizbuch zu finden, benötigt. Dies scheint zu zeigen, dass sich die funktionale Rolle von Ottos Notizbucheintrag von den entsprechenden Gehirnzuständen Ingas in einem relevanten Sinne unterscheidet. Zu diesem Thema siehe Clark und Chalmers (1998, S. 13f.) und Clark (2008, S. 80).

Die Kombination beider Thesen ist der erweiterte Funktionalismus. Aus diesem folgt das funktionale Paritätsprinzip, aus dem sich – mit Hilfe eines Beispiels – die These des erweiterten Geistes ableiten lassen könnte. Ein solcher Versuch scheitert jedoch am Fehlen einer funktionalistischen Theorie.

Auch der Versuch von Gegnern dieser These zu zeigen, dass ein zentrales Beispiel (das von Otto und seinem Notizbuch) die nötigen Bedingungen nicht erfüllt, scheitert, da eine – dem Psychofunktionalismus ähnliche – These benötigt wird. Eine solche ist jedoch kontraintuitiv. Eine Alternative stellt hier der Common-Sense Funktionalismus dar, der eine deutlich liberalere Auffassung von Geist und Kognition vertritt.

## Danksagungen

Zunächst möchte ich mich bei meinen Eltern Elke Hundertmark-Sagawe und Rainer Radloff bedanken, die mir dieses Studium ermöglicht haben. Ohne ihre Unterstützung wäre diese Arbeit wohl nicht zustande gekommen. Auch meinem Stiefvater Rainer Sagawe, der geholfen hat, so manchen Rechtschreibfehler aus meinen Essays und Hausarbeiten auszubügeln, will ich auf diesem Weg danken.

Weiterhin gilt mein Dank meinen Betreuern Dr. Holger Lyre und Dr. Peter Schulte und allen anderen, die es geschafft haben, mich für dieses Studium zu begeistern – besonders Prof. Dr. Eike von Savigny, Dr. Torsten Wilholt und mein Philosophielehrer Christian Kruggel seien hier erwähnt.

Zu guter letzt danke ich meinem Chef Dr. Hans-Jürgen Eikmeyer und allen anderen Menschen, deren Geduld ich in den letzten Monaten auf die Probe gestellt habe.



## Literatur

- [Adams und Aizawa 2008] ADAMS, Frederick ; AIZAWA, Kenneth: *The bounds of cognition*. Malden, Mass. [u.a.] : Blackwell, 2008
- [Beckermann 2000] BECKERMANN, Ansgar: *Analytische Einführung in die Philosophie des Gestes*. Berlin ; New York : Walter de Gruyter, 2000
- [Braddon-Mitchell und Jackson 1996] BRADDON-MITCHELL, David ; JACKSON, Frank: *Philosophy of Mind and Cognition*. Malden, Mass. [u.a.] : Blackwell, 1996
- [Chalmers 2008] CHALMERS, David: *Foreword*. S. IX–XVI. In: *Clark, Andy: Supersizing the Mind*, Oxford University Press, 2008
- [Clark 2005] CLARK, Andy: Intrinsic content, active memory and the extended mind. In: *Analysis* 65 (2005), Nr. 1, S. 1–11. – URL [www.era.lib.ed.ac.uk/bitstream/1842/1442/1/anal\\_514.pdf](http://www.era.lib.ed.ac.uk/bitstream/1842/1442/1/anal_514.pdf)
- [Clark 2008] CLARK, Andy: *Supersizing the Mind*. Oxford University Press, 2008
- [Clark und Chalmers 1998] CLARK, Andy ; CHALMERS, David: The Extended Mind. In: *Analysis* 58 (1998), Nr. 1, S. 7–19. – URL <http://consc.net/papers/extended.html>
- [Engelkamp und Zimmer 1994] ENGELKAMP, Johannes ; ZIMMER, Hubert D.: *The human memory*. Seattle, Washington [u.a.] : Hogrefe & Huber, 1994
- [Levin 2009] LEVIN, Janet: Functionalism. In: ZALTA, Edward N. (Hrsg.): *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab Center for the Study of Language and Information Stanford University, 2009. – URL <http://plato.stanford.edu/archives/sum2009/entries/functionalist/>
- [Lyre 2002] LYRE, Holger: *Informationstheorie*. München : Fink, 2002 (UTB ; 2289 : Philosophie)
- [Lyre 2008] LYRE, Holger: *Zum Extensionalismus des Mentalen*. 2008. – im Erscheinen
- [Müller 2003] MÜLLER, Horst M.: Neurobiologische Grundlagen der Sprachfähigkeit. In: RICKHEIT, Gert (Hrsg.): *Psycholinguistik*. de Gruyter, 2003 (Handbücher zur Sprach- und Kommunikationswissenschaft ; 24), S. 57–80
- [O’Brien 1998] O’BRIEN, Gerard: The Mind: Embodied, Embedded, but not Extended. In: *Metascience* 7 (1998), S. 78–83. – URL [http://cogprints.org/1451/0/Commentary\\_on\\_Clark\\_Being\\_There.pdf](http://cogprints.org/1451/0/Commentary_on_Clark_Being_There.pdf)
- [Rowlands 2003] ROWLANDS, Mark: *Externalism*. Montreal [u.a.] : McGill-Queen’s Univ. Pr., 2003

- [Rupert 2004] RUPERT, Robert D.: Challenges to the Hypothesis of Extended Cognition. In: *Journal of Philosophy* 101 (2004), Nr. 8, S. 389–428. – URL <http://spot.colorado.edu/~rupertr/ExtdRev5.pdf>
- [Shapiro 2008] SHAPIRO, Lawrence A.: Functionalism and Mental Boundaries. In: *Cognitive Systems Research* 9 (2008), S. 5–14. – URL <http://philosophy.wisc.edu/shapiro/HomePage/FMB.pdf>
- [Strube u. a. 1995] STRUBE, Gerhard ; HABEL, Christopher ; HEMFORTH, Barbara ; KONIECZNY, Lars ; BECKER, Barbara: Kognition. In: GÖRZ, Günther (Hrsg.): *Einführung in die künstliche Intelligenz*. Bonn [u.a.] : Addison-Wesley, 1995, S. 299–359
- [Turkington und Harris 2001] TURKINGTON, Carol ; HARRIS, Joseph R.: *The encyclopedia of memory and memory disorders*. New York, NY : Facts on File, 2001
- [Weiskopf 2007] WEISKOPF, Daniel A.: Patrolling the Mind’s Boundaries. In: *Erkenntnis* 68 (2007), Nr. 2, S. 265–276. – URL <http://shell.cas.usf.edu/~weiskopf/papers/patrolling-the-boundaries.pdf>
- [Wheeler im Erscheinen] WHEELER, Michael: *In Defense of Extended Functionalism*. In: MENARY, Richard (Hrsg.): *The Extended Mind*, Ashgate, im Erscheinen. – URL [http://www.philosophy.stir.ac.uk/staff/m-wheeler/documents/wheeler\\_extended\\_functionalism\\_submitted\\_revised\\_web.pdf](http://www.philosophy.stir.ac.uk/staff/m-wheeler/documents/wheeler_extended_functionalism_submitted_revised_web.pdf)

## Abbildungsverzeichnis

1	Eine Funktion mit den Eingangszuständen $E_1$ und $E_n$ und den Ausgangszuständen $A_1$ und $A_n$ . . . . .	4
2	Der schematische Aufbau des Colaautomatensystems. . . . .	6
3	Individuum-Umwelt-Interaktion (aus Müller, 2003, S. 59) . . . . .	12
4	Der schematische Aufbau des erweiterten Colaautomatensystems bei Anwendung des Paritätsprinzips. . . . .	19
5	Der schematische Aufbau des erweiterten Colaautomatensystems ohne Anwendung des Paritätsprinzips. . . . .	23