Philosophie des Geistes und Wissenschaftstheorie (1)[1]

Funktionalismus und die Reduzierbarkeit mentaler Phänomene

U. Mortensen

[1]Version Kognition3BNeuaab.tex, Erste Fassung: 14. 04. 2023, überarbeitete Fassung: 03. 12. 2024, Kognition3BHilfe.pdf 16. 02. 2025, letzte Fassung 26. 02, 2025 KognitionPhilMind(I)Funktion.tex

Inhaltsverzeichnis

1 Behaviorismus, Kognitivismus, Physikalismus

Bis in die 1950er Jahre war der von James B. Watson (1878 – 1958) konzipierte Behaviorismus (Watson (1913)) vor allem in der anglo-amerikanischen, aber auch in Teilen der deutschen akademischen Psychologie dominierend: die Grundannahme dieses Ansatzes ist, dass das Verhalten von Personen auf der Basis von Verhaltensdispositionen erklärt werden kann, ohne dass dabei auf nicht direkt beobachtbare innere bzw. mentale Zustände zurückgegriffen werden muß; psychologische Theorien werden gerne als *Stimulus-Response-Theorien* formuliert. Mit der in den fünfziger Jahren einsetzenden 'kognitiven Wende' werden in zunehmendem Maße auch mentale Prozesse Gegenstand psychologischer Forschung, wobei es natürlich erscheint, dass die Aktivität des Gehirns die Basis mentaler Prozesse ist; Place (1956) und Smart (1959) publizieren *Identitätstheorien*, in denen postuliert wird, dass mentalen Ereignissen oder Phänomenen spezifische neuronale Prozesse zugeordnet werden können. Damit wird unterstellt, dass mentale Ereignisse auf letzlich physikalische Prozesse zurückgeführt werden können. Diese Annahme charakterisiert den *Physikalismus*, womit der klassische Materialismus gemeint ist, allerdings ohne dass metaphysische Annahmen wie die Nichtexistenz Gottes mitgedacht werden: es geht nur um die Rückführbarkeit mentaler auf materielle, also physikalische und biochemische Prozesse. In der Philosophie des Geistes – wobei der Begriff 'Geist' dem des englischen *mind* entspricht, Begriffe wie etwa der objektive und subjektive Geist der hegelschen Philosophie werden also nicht mitgedacht – wird Rückführbarkeit oft als *Reduktion* bezeichent; der 'Reduktivismus' bezeichnet also die Theorie, dass mentale auf physische Prozesse "reduziert" werden können. Während der reduktivistische Ansatz in der Neuro- und Kognitionswissenschaft gewissermaßen als Hintergrundstheorie mitläuft findet man in der Philosophie des Geistes seit den sechziger Jahren des 20-ten Jahrhunderts dualistische oder quasi-dualistische Ansätze wie den *nonreductive physicalism*, die ihren Hintegrund weniger in klassischen metaphysischen Betrachtungen haben, sondern letztich in metamathematischen Betrachtungen zur Beweisbarkeit mathematischer Theoreme.

Bevor darauf näher eingegangen wird, soll noch ein kurzer Blick auf den behavioristischen Stimulus-Response-Ansatz geworfen werden. Das Verhalten löst Reaktionen aus der Umwelt aus, die die jeweiligen Dispositionen verstärken oder hemmen. Den Ausdruck 'Disposition' kann man durch 'bedingte Wahrscheinlichkeit' ersetzen und dafür $P(R|S)$ schreiben, wobei P die Wahrscheinlichkeit der Reaktion R ist unter der Bedingung, dass der Stimulus durch S gegeben ist. Die philosophische Forderung hinter dieser Annahme ist, dass gesetzartige Beziehungen sich nur auf Beobachtbares beziehen dürfen, – das nicht direkt Beobachtbare ist "metaphysisch", wobei dieses Prädikat abwwertend gemeint ist. Diese Forderung spielte auch in der Frühphase des Neopositivismus bzw. des logischen Empirismus des Wiener Kreises eine zentrale Rolle. Burrhus

Frederic Skinner (1904 – 1990), ein namhafter Vertreter des Behaviorismus, publizierte 1957 das Buch *Verbal Behavior*, in dem das Lernen von Sprache wie anderes Lernen auch durch Verstärkungs- und Hemmungsprozesse (operantes Konditionieren) erklärt wird. Ein expliziter Bezug auf das Bewußtsein oder "kognitive" Prozesse gilt im Rahmen des Behaviorismus als eher unwissenschaftlich, weil diese Prozesse im Gegensatz zu Reiz-Reaktions-(Stimulus-Response)-Paaren nicht direkt beobachtbar sind. Der Charme des behavioristischen Ansatzes bestand zu einem nicht geringen Teil darin, das vage Ausdrücke aus dem Bereich der idealistischen Philosophie[2] wie etwa "der lebendige Geist" (Dilthey) nicht vorkamen und komplexe psychische Prozesse durch geeignet definierte Stimulus(S)-Response(R)-Kombinationen beschreibbar gemacht werden sollten. Allerdings wurde immer deutlicher, dass die Ausklammerung der mentalen Zustände, die zwischen S und R vermitteln zu einer gewissen Sterilität der Forschung führte. Der Versuch, die Wissenschaft metaphysikfrei begründen zu wollen, ist selbt ein metaphysischer Ansatz, wie die Philosophen des Wiener Kreises einsahen. Warum sollen auch natürlich erscheinende Fragen wie die, wie das Gehirn aus neuronalen Kodierungen elektromagnetischer Wellen bestimmter Länge die subjektive Erfahrung "rot" erzeugt, oder wie neuronale Teilpopulationen logische Operationen oder emotionale Dynamiken generieren als unwissenschaftlich gelten? Die Zeit war reif geworden für einen Angriff auf den Behaviorismus: 1959 erschien Noam Chomskys "A Review of B. F. Skinner's Verbal Behavior"(Chomsky (1959), in der er den Behaviorismus am Beispiel des Skinnerschen Ansatzes einer grundsätzlichen Kritik unterzog. Ein wesentlicher Punkt dieser Kritik ist, dass interne, nicht direkt beobachtbare mentale Zustände wieder explizit in der Kognitionsforschung berücksichtigt werden sollten[3]. Chomskys Besprechung des Skinnerschen Werkes gilt als Beginn des *Kognitivismus*.

Mentale Zustände sind im Allgemeinen nicht direkt beobachtbar und müssen dementsprechend erschlossen werden. Dazu muß definiert werden, was mit dem Ausdruck 'mental' gemeint ist. Den gängigen Nachschlagewerken zufolge steht er abkürzend für 'kognitiv', d.h. für den Effekt von Wahrnehmungen, von gedanklichen Prozessen, aber auch von Stimmungen etc. Im Deutschen steht 'mental' auch für 'geistig', wobei dieses Prädikat auch immateriell oder spirituell gedachte Aspekte bewußter Zustände meinen kann. Eine genauere Bestimmung des Begriffs ergibt sich, wenn man der Frage nachgeht, ob 'mental' äquivalent mit 'bewußt' ist. Beckermann (2008), p. 9, listet einige Charakteristika mentaler Ereignisse auf, wie sie in der Philosophie des Geistes diskutiert

[2]Gemeint ist der die Philosophie des Idealismus, wie sie etwa von Immanuel Kant vertreten wurde.

[3]Dass Chomsky dabei seinerseits über das Ziel nüchterner Kritik hinausschoss und eher intuitive Ansichten als zwingend darstellte wurde im Überschwang der Befreiung vom behavioristischen Joch kaum bemerkt. Erst 1970 erschien MacCorquodales Kritik an der Chomskyschen Kritik, in der er ausführt, warum der Skinnersche Ansatz nicht *in toto* verworfen werden kann.

werden:

1. Mentale Phänomene unterscheiden sich von physischen Phänomenen dadurch, dass sie bewußt sind,
2. Mentale Zustände unterscheiden sich von physischen Phänomenen, da unser Wissen um unsere eigenen mentalen Zustände unkorrigierbar ist,
3. Mentale Phänomene unterscheiden sich von physischen durch ihre Intentionalität.

Darüber hinaus wird noch die Nicht-Räumlichkeit und die Privatheit mentaler Ereignisse genannt. Die Bewußtheit eines mentalen Ereignisses ist sicher ein zentrales Merkmal, sie bedeutet, dass eine Person, die in einem mentalen Zustand M ist, *weiß*, dass sie in diesem Zustand ist. Beckermann verweist aber darauf, dass dieses Merkmal weniger klar ist, als es auf den ersten Blick scheint. So ist 'unter psychischem Stress stehen' sicher ein mentaler Zustand, aber es fragt sich, ob man sich wirklich immer auch bewußt ist, dass man unter Stress steht. Die Frage verweist auf eine Art von Selbstreferentialität des Bewußtseins: es kann mentale Zustände über mentale Zustände geben. Daraus ergibt sich die Frage, ob es möglich ist, mit rein philosophischen Mitteln herauszufinden die Genese eines zumindest selbstreferentiellen Bewußtseins zu begreifen, und wenn nicht, ob diese Genese empirisch, mit naturwissenschaftlichen mitteln zu decodieren ist. Natürlich kann bereits diese Frage als eine philosophische Frage betrachtet werden, die ihre eigenen selbstreferentiellen Aspekte hat. Um sich nicht darin zu verlieren könnte es am gesündesten sein, sich ohne weitere Rechtfertigung auf die Empirie zu konzentrieren: man wird sehen, wie weit man kommt, und ein paar Einsichten wird man auf jeden Fall erhalten.

Die Annahme, dass die oben in Punkt 3. erwähnte Intentionalität ein Merkmal des Geistigen sei hat in der Philosophie eine lange Tradition, die Franz Brentano (1838 – 1917) in der neueren Philosophie wiederbelebt hat. Allerdings wird sie keineswegs von allen Kognitions- und Neurowissenschaftlern geteilt, denn es fehlt schlicht ein Argument, aus dem hervorgeht, warum die Gerichtetheit des Bewußtseins die Nicht-Physikalität des Bewußtseins belegen soll. Dafür gibt es neurowissenschaftliche Untersuchungen, die Aufschluß über die neuroanatomische und neurophysiologische Basis der Intentionalität liefern, etwa Zapparoli et al. (2017)(weitere Literatur dort). Vielleicht ist es forschungslogisch gesehen gar nicht besonders vernünftig, von vorn herein und abstrakt nach Charakteristika des Mentalen zu suchen, weil man sich in einem unproduktiven Streit um Grundsätzlichkeiten verfängt. Einige Positionen seien aber gleichwohl genannt, weil sie immer wieder in den Argumentationen auftauchen. So hat insbesondere die Frage nach der Natur der Qualia die Diskussion um das Leib-Seele-Problem bzw. den Dualismus wieder aufflackern lassen. Es haben sich sich verschiedene philosophische Schulen entwickelt: es gibt die die Materialisten – auch Naturalisten oder Physikalisten genannt – die von einer Rückführbarkeit mentaler auf neuronale, also letzlich physikali-

sche Prozesse ausgehen, gemäßigte Naturalisten, die zwar annehmen, dass das Mentale nicht in einer von der Materie separierten *res cogitans* verortet ist, die aber nicht an eine Reduzierbarkeit mentaler Prozesse auf bestimmte neuronale Abläufe glauben (*Quasi-* oder *Eigenschaftsdualismus*) (Beckermann (2008), p. 7), und schließlich die Dualisten, die einfach postulieren, das Mentale habe keine materielle Basis, – ein philosophischer Ansatz, den man schon bei Platon findet. Will man nicht abstreiten, dass neuronale Aktivität eine notwendige Rolle für mentale Prozesse spielt, so kann man postulieren, dass das Mentale ein emergentes, irgendwie nichtphysisches Phänomen ist. Wenn aber neuronale Aktivität als einerseits notwendige, andererseits aber nicht hinreichende Bedingung gedacht wird, so wird man erklären müssen, wie diese Emergenz des Nichtphysischen entsteht. Konstruktive Überlegungen zur Genese von Emergenz scheinen in der Philosophie des Geistes noch auszustehen, wobei es denkbar ist, dass die Natur der Emergenz nicht mit rein philosophischen Mitteln aufklärbar ist.

Funktionalismus: Im Folgenden soll insbesondere der zuerst von Hilary Putnam (1926 – 2016) in den Jahren (1960) und (1967) konzeptualisierte *Funktionalismus* diskutiert werden, weil dieser zu den einflußreichsten philosophischen Betrachtungen zur Kognitionsforschung gehört, und andererseits zu langen und wichtigen Kontroversen u.a. über den Status der Psychologie in Bezug auf die Neurobiologie geführt hat. Putnam knüpfte dabei an Überlegungen an, die scbon von dem britischen Mathematiker Alan Turing (1950) angestellt worden waren, in dem er die Frage stellte, ob es zumindest im Prinzip möglich sei, das Gehirn als einen Computer zu modellieren. Die Frage führt zu grundsätzlichen Überlegungen über mögliche Grenzen der Erklärbarkeit psychischen Geschehens, wie sie von Lucas (1960) vor dem Hintergrund der Gödelschen Theoreme gestellt wurden. Der österreichische Mathematiker und Logiker Kurt Gödel[4] hatte 1931 gezeigt, dass es in axiomatisierten, konsistenten mathematischen Theorien Theoreme gibt, die sich im Rahmen dieser Theorie nicht beweisen lassen, d.h. die sich nicht auf die gegebenen Axiome zurückführen lassen, und Lucas argumentierte auf der Basis dieses Befundes, dass das menschliche Gehirn jedem programmgesteuerten Apparat überlegen sei, – und mithin nicht durch Strukturen erklärbar sei, die als Computerprogramme definiert seien. Lucas' Argumente sind bis heute umstritten, worauf kurz eingegangen werden wird.

2 Funktionalismus und Komputationalismus

Der Begriff des Funktionalismus ist allgemeiner als der des komputationalen Funktionalismus, so dass zunächst ein paar allgemeine Aussagen zum Funktionalismus gemacht werden:

[4] 1906 – 1978

> ... in the philosophy of mind is the doctrine that what makes something a mental state of a particular type does not depend on its internal constitution, but rather on the way it functions, or the role it plays, in the system of which it is a part. (Levin (2018))

Wie Levin weiter ausführt liegen die Wurzeln des Funktionalismus in der Antike, insbesondere bei Aristoteles, der um 350 v.Chr. unter anderem Betrachtungen über die menschliche Seele anstellte. Diese mache die essentielle Qualität (die *quidditas*, also die Washeit oder "das Wesen") des Menschen aus, die die Funktion bzw. den Zweck eines Menschen definiere. Es war der amerikanische Mediziner, Philosoph und Psychologe William James (1847 – 1010), der die Idee des Funktionalismus in der Psychologie wieder aufnahm, indem er mentale Zustände als Funktionen aus einer Gesamtheit von Zuständen beschrieb, die der Anpassung eines Organismus an seine Umwelt dienten (Arnold et al (1980), p. 651). Der Jamessche Funktionalismus wird allerdings in den folgenden Abschnitten nicht Gegenstand der Betrachtungen sein, sondern die durch Hilary Putnam (1926 – 2016) formulierte Variante (Putnam (1960)), die er 1967 zu einem *komputationalen Funktionalismus* spezifizierte, der in den folgenden Jahrzehnten in der Philosophie des Geistes eine dominierende Rolle spielte.

2.1 Mentale Zustände und Definition des Funktionalismus

Der noch bis in die 60er Jahre des zwanzigsten Jahrhunderts in der empirischen Psychologie dominierende Behaviorismus war spätestens mit Noam Chomskys (1959) Besprechung des Buchs *Verbal Behavior* von Burhus Frederic Skinners (1904 – 1990) in die Kritik geraten; dem behaviouristischen Credo zufolge galt es als unwissenschaftlich, in psychologischen Aussagen vom Bewußtsein zu sprechen. Putnam begann mit einer 1960 und 1967 einsetzenden Folge von Arbeiten mit der Konzeption des Funktionalismus, der im Prinzip auf einer Definition des Begriffs des mentalen Zustands beruht. Nach Blocks (1996) durch viele Diskussionen gereifte Definition soll gelten:

> Geistige Zustände werden nur durch ihre funktionale Rolle konstituiert, d.h. durch ihre kausalen Beziehungen zu anderen mentalen Zuständen, zu sensorischen Inputs und Verhaltensweisen[5].

Der Punkt bei dieser Definition ist offensichtlich, dass sie auf die *quidditas*, die "Washeit" von mentalen Zuständen gar nicht eingeht: damit ist das phänomenale Erleben von Empfindungen wie Schmerz oder der Farbe eines Objekts, also von Qualia, gemeint. Levin (2018) schreibt dann auch

[5] [mental states] are constituted solely by their functional role, which means, their causal relations with other mental states, sensory inputs and behavioral outputs.

> That is, what makes something a mental state is more a matter of what it does, not what it is made of.

Der Reiz der funktionalistischen Definition von mentalen Zuständen besteht darin, gar nicht erst auf die Frage nach der internen Konstitution eines mentalen Zustands einzugehen: Es wird nicht auf die Frage nach der *quidditas*, also das, was Zustandsbeschreibungen wie "traurig sein", "verärgert sein", "frustriert" bzw "aggressiv sein" bedeuten, eingegangen, vielmehr soll sich die Bedeutung durch ihre Funktion in einem funktionalen Netzwerk von Zuständen erklärt werden; man könnte diese Art der Definition von Zuständen "nicht-essentialistisch" nennen. Solche Definitionen erinnern an den Begriff der operationalen Definition (Bridgman (1927)) von Begriffen, wie sie 1927 von dem amerikanischen Physiker Percy Williams Bridgman (1882 – 1961) in die Physik eingeführt wurde: Begriffe wie 'Aggression', 'Frustration' werden nicht über die Bedeutung dieser Wörter erklärt, sondern über die Methode der Messung der Intensität dieser Zustände. Putnam hat allerdings den Operationalismus kritisiert[6] weil er dem tatsächlichen Gebrauch der Begriffe nicht entspräche, – eine Kritik, die allerdings auch auf seine Definition des Begriffs des mentalen Zustands anwendbar ist.

Der Putnamsche Ansatz wurde nicht zuletzt auch durch die Arbeiten von Putnams Schüler Jerry Fodor (1935 – 2017) zum *received view*[7], also gewissermaßen zu einer allgemein anerkannten Standardtheorie im Rahmen der Philosophie des Geistes (Shagrir (2005)).

Die funktionalistische Definition des mentalen Zustands hat, weil sie ohne den Bezug auf die interne Konstitution des jeweils Mentalen auszukommen versucht, einen gewissermaßen rekursiven Charakter – ein mentaler Zustand entsteht durch Interaktion von mentalen Zuständen, die ihrerseits als Interaktion von mentalen Zuständen definiert sind, etc. Tatsächlich ist es ja so, dass ein Zustand einen anderen auslösen kann: gerät eine Person in den Zustand des Frustriertseins, so *kann* sie als Folge davon in einen Zustand der Aggressivität geraten, – abhängig von anderen Aspekten ihres Gesamtzustands. Andererseits ergibt sich die Frage, wann und ob die Rekursion irgendwann endet, d.h. ob implizit angenommen wird, dass Basiszustände existieren, die nicht weiter als Interaktion anderer Zustände erklärt werden können.

Der Funktionalismus ist hinsichtlich vieler Details diskutiert worden, über die man sich in verschiedenen Übersichtsartikeln orientieren kann (z. B. Levin (2023), Polger (IEP)). So werden drei Typen von Funktionalismus voneinan-

[6] "It is well known that narrow operationalism cannot successfully account for the actual use of scientific or common-sense terms.", in: The Meaning of "Meaning". In: Putnam, H.: Mind, Language and Reality, Philosophical Papers, Volume 2. Cambridge University Press 1975

[7] Dieser Ausdruck bezeichnet wissenschaftliche Theorien als "axiomatic calculi in which theoretical terms are given a partial observation interpretation by means of correspondence rules", Suppe (1989), p. 38 .

der unterschieden: 1. Psycho-Funktionalismus, 2. Analytischer Funktionalismus, und 3. Komputationaler Funktionalismus. Der Psycho-Funktionalismus basiert auf der Ablehnung des Behaviorismus und zielt auf eine Grundlegung einer empirisch orientierten Kognitionswissenschaft. Der Analytische Funktionalismus soll die Beziehungen zwischen den kausalen Relationen zwischen mentalen Zuständen zu (a) Stimulierungen, (b) Verhaltensweisen und (c) anderen mentalen Zuständen klären (Dollard & Miller (1939)). Weshalb allerdings Philosophen meinen, empirisch arbeitenden Psychologen eine Rahmenphilosophie liefern zu müssen (Searle (1992)) "Psychology needs a philosophical foundation") ist unklar. Sicher haben Philosophien in der Entwicklung der Wissenschaften eine Rolle gespielt, aber nicht notwendig eine konstruktive: es waren Tycho Brahes Daten und deren Analyse durch Kepler, die die philosophisch-theologisch begründete Kreisbahntheorie für Planeten überwanden, es war andererseits Ernst Mach, der aufgrund seiner positivistischen, also philosophischen Überzeugung Ludwig Boltzmanns atomistischen Ansatz bei der Entwicklung der Thermodynamik kritisierte (und behinderte): "Haben'S schon mal eins g'sehen?" soll er Boltzman gefragt haben (Scheibe (2006)), – gemeint war das Atom, an dessen Existenz er nicht glaubte. Und der Philosoph Alva Noë hat herausgefunden, dass weder Bewußtsein noch das Handeln vom Gehirn abhängen, was auch Neurowissenschaftler einsehen könnten, wenn sie sich nur um die Entwicklungen der lezten sechzig Jahre in der Philosophie kümmern würden, – im Kapitel über den freien Willen wird näher darauf eingegangen.

Durchgesetzt haben sich am Ende die Ideen, die in enger Wechselwirkung zwischen der Entwicklung theoretischer Vorstellungen, also der Formulierung von Hypothesen, und der empirischen Überprüfung der Hypothesen entstanden.

Die dritte Variante des Funktionalismus, der von Putnam 1967 eingeführte komputationale Funktionalismus hat seine Basis weniger in der Kognitions- oder neurowissenschaftlichen Forschung, sondern in der Mathematik, genauer: in der Metamathemtik. Sowohl Putnam wie Fodor haben in ihren Arbeiten nach 1967 unter Funktionalismus stets den komputationalen Funktionalismus verstanden. Im Folgenden wird auf diese Variante des Funktionalismmus fokussiert.

2.2 Komputationaler Funktionalismus

2.2.1 Turings Erkenntnisse

Es begann mit einer Krise der Mathematik in den letzten Jahrzehnten des 19ten Jahrhunderts, - es ging um die Paradoxien des Unendlichkeitsbegriffs, um den Begriff der Zahl und um die Frage, wie eine widerspruchsfreie Ma-

thematik erreicht werden könnte. Im Jahr 1899 veröffentlichte David Hilbert[8] sein Werk *Grundlagen der Geometrie*, in dem er ein vollständiges Axiomensystem der euklidischen Geometrie, das im Unterschied zu den euklidischen Axiomen nicht mehr auf die Anschauung zurückgriff und die Geometrie als rein formales System aufzubauen gestattete. Der Vorteil dieses Ansatzes besteht darin, dass er frei von den Mehrdeutigkeiten der Anschauung ist. In den 20er 'Jahren de 20ten Jahrhunderts entwickelte Hilbert das dann nach ihm benannte *Hilbertprogramm*, ein Forschungsprogramm, in dem mit *finiten* Methoden die Widerspruchsfreiheit der Mathematik bewiesen werden sollte. Es sollte ein Axiomensystem gefunden werden, das (i) widerspruchsfrei, also konsistent, und (ii) vollständig sein sollte, d.h. jeder wahre Satz sollte aus dem System ableitbar sein. Idealerweise sollte ein Algorithmus – gemeint war zunächst eine systtematische Folge von Operationen – existieren, dem eine Aussage eingegeben werden kann und der dann nach einer endlichen Folge von Schritten entscheidet, ob die Aussage aus den jeweiligen Axiomen abgeleitet werden kann oder nicht. Die Frage nach der Existenz eines solchen Algorithmus wurde von Hilbert & Ackermann (1928) das *Entscheidungsproblem* genannt (dieser Ausdruck wird auch in englischsprachigen Texten verwendet).

Aber 1930/1931[9] publizierte Kurt Gödel seinen *Unvollständigkeitssatz*, der eigentlich aus zwei Theoremen besteht:

> **1.** F sei ein hinreichend reichhaltiges formales System. Ist F konsistent, dann ist F unvollständig.
> **2.** F sei konsistentes formales System. Dann kann mit den Mitteln von F nicht gezeigt werden, dass F konsistent ist. Zitat 1

"Unvollständig" heißt dabei, dass nicht *jeder* wahre Satz bewiesen, d.h. aus den Axiomen hergeleitet werden. "hinreichend reichhaltig" heißt, dass z.B. die bekannten Eigenschaften der natürlichen Zahlen in F definiert werden, also bei Beweisen vorausgesetzt werden können[10]. Die beiden Sätze bedeuteten das Ende des Hilbertprogramms. Im Rahmen der Philosophie des Geistes bedeuten sie, so einige Autoren, die grundlegende Inadäquatheit bestimmter Theorien. Auf Gödels Beweis seines Theorems kann hier nicht eingegangen werden, er geht weit über den Rahmen dieses Textes hinaus. Es sei aber angemerkt, was seinen Kern ausmacht. Er geht auf eine schon im Altertum bekannte Problematik der Selbstreferentialität zurück, nämlich die Paradoxa des Eubulides (viertes Jahrhundert vor Chr.).Das Bekannteste ist das Lügner-Paradoxon: ein Kreter sagt, "Alle Kreter lügen". Sagt der Kreter die Wahrheit, so lügt

[8] 1863 – 1943

[9] zuerst auf der "2. Tagung für Erkenntnislehre der exakten Wissenschafen mit dem Tema *Grundlagen der Mathematik* vom 5.–7. Sep. 1930 in Königsberg", dann in einem Aufsatz "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I" in den *Monatshefte für Mathematik und Physik,* 38, 1931, S. 173–198

[10] Eine sehr gute Darstellung der Gödelschen Resultate findet man in Hoffmann (2011)

er, denn das ist die Aussage des Satzes. Lügt er aber, so ist der Satz also wahr, so dass er lügt, etc. In allgemeiner Form kann man das Paradoxon so formulieren: A sei der Satz "Der Satz A ist falsch". Ist A wahr, dann ist A falsch, und ist A falsch, so ist A wahr, etc. Das Paradox ergibt sich aus der Selbstreferenz des Satzes. Gödel formulierte eine Variante des Paradoxes, die in der Formulierung von Hoffmann (2011), p. 201), zur Basis seines Beweises wird:

(*) "Ich bin innerhalb des Kalküls nicht beweisbar." Zitat 2

Mit Kalkül ist ein formales System gemeint. Weder eine derartige Aussage noch deren Negation ist unter den Voraussetzungen des Gödelschen Theorems aus einem gegebenen Axiomensystem ableitbar, weshalb der Satz (*) unentscheidbar ist.

1936 publizierte der britische Mathematiker Allan Turing einen Aufsatz über die Berechenbarkeit von Zahlen. Was diese Arbeit zu einem Meilenstein der Mathematik und Logik machte war Turings Charakterisierung des Begriffs des Algorithmus. Bereits Hilbert hatte gefordert, für bestimmte Fragestellungen – in seinem Fall das Bestimmen von Lösungen von *diophantischen Gleichungen* ein Verfahren anzugeben, "nach welchem sich mittels einer endlichen Anzahl von Operationen entscheiden lässt, ob die Gleichung in ganzen rationalen Zahlen lösbar ist"[11]. Ein Beispiel für eine solche Gleichung ist das "Pythagoreische Tripel" $a^2 + b^2 = c^2$ mit $a, b, c \in \mathbb{N}$, $\mathbb{N}$ die Menge der natürlichen Zahlen; ein bekanntes Beispiel sind die Zahlen $a = 3, b = 4, c = 5$. Ebenso kann man nach einem systematischen Verfahren fragen, das Formeln der Prädikatenlogik erster Stufe[12] auf ihre Korrektheit prüft, oder sie herleitet. Einen analoge Frage kann man für Beweise arithmetischer Formeln stellen. Es geht zunächst nicht darum, wie ein Mathematiker einen solchen Beweis findet, sondern darum, dass der Beweis als eine bestimmte Folge von Operationen dargestellt werden kann; diese Folge zeigt die Beweisbarkeit einer Aussage. Der Beweis ist eine *endliche* Folge von Operationen, an deren Ende die Entscheidung "ist wahr" oder "ist nicht wahr" steht. In diesem Sinne ist eine Aussage beweisbar, wenn sie entscheidbar ist. Man kann sich ein Verfahren vorstellen, dessen Anwendung eben diese Folge generiert. Ein derartiges Verfahren ist ein Algorithmus. Gesucht ist eine eindeutige Charakterisierung dessen, was einen Algorithmus ausmacht. Turing (1936) lieferte sie in Form einer Beschreibung einer von ihm a-Maschine genannte Maschine konzipierte, (a für automatic), und für die später der Ausdruck 'Turingmaschine' üblich wurde. Eine derartige Maschine repräsentiert ein Programm, das die Ausführung eines mechanischen Prozess bewirkt. die Maschine besteht aus einem Lesekopf, durch den in diskreten Schritten ein Papierstreifen läuft. Auf dem Streifen ist eine Folge von

[11] s. Hoffmann (2011), p. 236

[12] Das sind Aussagen der Form "für alle x, wenn x ein Mensch ist, dann ist x sterblich".

Kästchen gezeichnet. In jedem Kästchen ist ein Symbol aus einer endlichen Menge von Symbolen gezeichnet. Der Lesekopf liest das Symbol und ersetzt es nach Maßgabe des Programmes, dass die Operationen des Lesekopfes steuert. Auf diese Weise wird eine Folge von Symbolen in eine andere Folge überführt. Insbesondere können auch Zahlen als Folgen von Symbolen, z.N: Nullen und Einsen, dargestellt werden. Die Transformation einer Symbolfolge läuft dann auf die Berechnung einer Zahl hinaus; eine ausführliche Darstellung findet man in Hoffmann, p. 53. Dies erklärt den Ausdruck 'Berechnung' und 'Berechenbarkeit', auch wenn Folgen von Aussagen erzeugt werden. Die Turingmaschine erweist sich als ein Modell für die Lösung aller Aufgaben, für die nach endlich vielen Schritten eine Lösung gefunden werden kann. Ein Problem wird demnach nicht gelöst, wenn die Maschine nicht nach einem bestimmten Schritt hält. Wenn sie nicht hält, existiert keine Lösung, das Problem ist nicht entscheidbar. Natürlich kann es sein, dass das zu lösende Problem sehr komplex ist und sehr lange läuft. Man möchte gerne wissen, ob sie irgendwann stoppen wird, was zu der Idee führt, nach einer Turingmaschine zu fragen, die entscheiden kann, ob die in Frage stehende Maschine irgendwann hält; dies ist das *Halteproblem*, und Turing konnte 1936 zeigen, dass es keinen Algorithmus gibt, der die Frage nach dem möglichen Stoppen für alle mögliche Algorithmen und alle möglichen Eingabebänder beantworten kann. Man sagt, das Halteproblem sei algorithmisch nicht entscheidbar.

Man stelle sich nun vor, dass eine Turingmaschine T_a den Beweis einer Aussage liefern soll. Stoppt die Maschine nicht nach endlich vielen Schritten, so kann sie keinen Beweis finden, und dem Halteproblem zufolge existiert auch kein Algorithmus, der entscheiden könnte, ob T_a jemals stoppt. Turing hat also mit dem Halteproblem einen dem Gödelschen Unvollständigkeitssatz analogen Satz bewiesen. Das Hilbert-Ackermannsche Entscheidungsproblem hat also keine Lösung[13].

Außer Turing beschäftigte sich auch der amerikanische Logiker Alonzo Church[14] mit der Frage der Berechenbarkeit (im hier gemeinten allgemeinen Sinn) und entwickelte den λ-calculus, von dem gezeigt werden kann, dass er dem Begriff der Turingmaschine äquivalent ist; so kam es zur **Church-Turing-These**:

> **Church-Turing:** Die Klasse der Turing-berechenbaren Funktionen stimmt mit der Klasse der intuitiv berechenbaren Funktionen überein.

(Hoffmann (2011), p. 263). Diese These ist nicht beweisbar, weil der Begriff der 'intuitiv berechenbaren Funktion' nicht scharf definiert ist. Gemeint sind Funktionen, die von Menschen, auf welche Weise auch immer, berechnet wer-

[13] Hinweis auf Leibniz' machina rationatrix?

[14] 1903 – 1995

den können. Demnach besagt die Church-Turing-These, dass alles, was von Menschen überhaupt berechnet werden kann, durch eine Turingmschine berechnet werden kann (Hoffmann, p. 253). Turing (1950) vermutete dann, dass jede kognitive Aktivität eines Menschen, d.h. seines Gehirns, in der Sprache von Turingmaschinen formuliert werden kann.

2.2.2 Komputationaler Funktionalismus

Der Neurophysiologe Warren McCulloch und der Logiker Walter Pitts waren wohl die Ersten, die ein Modell des Zuammenwirkens von Neuronen vorstellten, um logische Operationen zu erzeugen (McCulloch & Pitts (1943)). Es war stark vereinfachtes Modell, es ging nur um prinzipielle Annahmen über die Interaktion von Neuronen, die auf logische Schlußfolgerungen hinauslaufen. Wenn dabei von 'computation', also von Berechnung die Rede war, so haben sie nicht an einen Computer heutiger Bauart gedacht, denn die Architektur heutiger Rechner wurde erst 1945 von John von Neumann vorgeschlagen. Sie werden eher an ein algorithmisch arbeitendes "Gerät"' gedacht haben, das neuronale Netze und damit biologische Organismen in approximierter Form abbildet. Sie waren unter anderem der Ansicht, durch Konstruktion neuronaler Netze das Leib-Seele-Problem lösen zu können. Einige bedeutende Mathematiker wie Norbert Wiener (1894 – 1964)[15] und der Mathematiker John von Neumann (1903 – 1957), der u.a. die Architektur moderner Computer und dabei die Verwendung von Flußdiagrammen zur Beschreibung des Aufbaus von Computern und Programmnen entwickelte, gingen einfach davon aus, dass eine derartige Repräsentation möglich sei. John von Neumann interpretierte McCulloch and Pitts's Arbeit als Beweis, dass

> "anything that can be exhaustively and unambiguously described, anything that can be exhaustively and unambiguously put into words, is ipso facto realizable by a suitable finite neural network."

vom McCulloch-Pitts Typ (von Neumann 1951, 23.) Künstliche neuronale Netze wie die von McCulloch & Pitts konzipierten lassen sich auf einem Computer simulieren, so dass the These, menschliche Kognitionen ließen sich alle auf dem Computer simulieren schnell in die Gedankenwelt der an der Natur von Kognitionen interessierten Philosophen und in der Folge auch in die der Psychologen gelangte. Die Turingmaschine galt als generisches Beispiel für einen Computer. Bemerkenswert bei diesen Ansätzen ist, dass bei den Überlegungen von McCulloch, Pitts und Wiener die Frage nach dem Bewußtsein gar nicht angesprochen wird. Es geht einfach nur darum, grundlegende Mechanismen zu

[15] N. Wiener wurde der wissenschaftlich Interessierten Allgemeinheit durch sein zuerst 1948 und dann 1961 in verbesserter Form erschienenes Buch *Cybernetics* mit dem Untertitel 'on control and communication in the animal and the machine' bekannt, in dem u.a. die Beziehung zwischen Computern und dem Nervensystem diskutiert wurde.

modellieren, um zu ersten Vorstellungen über die Wirkungsweise neuronaler Systeme zu gelangen.

1958 publizierten Paul Oppenheim[16] und Hilary Putnam[17] (1958), die These, dass Turings Analyse des Begriffs der Berechenbarkeit[18] führe "natürlich" zu der These, dass das Gehirn eine Turingmaschine sei (s. Piccinini (2004) p. 813); das McCulloch & Pitts-Modell kann als die spezielle Realisierung einer Turingmaschhine gesehen werden. In den empirischen Wissenschaften würde man von der Oppenheim-Putnamschen These eher als von einer *Hypo*these sprechen, die geprüft werden muß, aber die Frage nach einer empirischen Überprüfung spielte bei diesen Überlegungen gar keine Rolle. Der *komputationale* Funktionalismus wurde von Putnam 1967 zunächst als Gegenthese zum Behaviorismus und zur Identitätstheorie ("type-identity theory")[19] von Place (1956) und Small (1959) eingeführt. Die große Beachtung, die der komputationale Funktionalismus insbesondere bei Philosophen erfuhr scheint sich aus Putnams Feststellung zu ergeben. dass die funktionale Organisation etwa eines Menschen als Folge von mentalen oder logischen Zuständen ohne Referenz zur physischen Realisierung dieser Zustände beschrieben werden kann. Die Dynamik der Zustandsübergänge sei im Prinzip durch einen Automaten mit probabilistischen[20] Zustandsübergängen beschreibbar. Putnam führt aus, dass das kognitive System des Menschen als ein komputationales System analog zu einem Computer konzipiert werden könne, ohne allerdings die Beziehungen zwischen einer Repräsentation in einem Computer und der in einem neuronalen System zu elaborieren.

Eine Turingmaschine als allgemeines Modell einer funktionalistischen Dynamik bedeutet allerdings, dass die mentalen Zustände durch eine abzählbare, also diskrete Menge $S_1, S_2, \ldots, S_n, \ldots$ gegeben sind, deren Interaktion durch

$$S_i \xrightarrow{o_k} S_j, \quad i, j = 1, \ldots, n, \quad i \neq j \tag{1}$$

angegeben werden kann, wobei o_k eine auf S_i angewendete Operation ist, die den Übergang von S_i zu S_j bewirkt. Das Speicherband kann im Prinzip unendlich lang sein. Da ein Programm stets durch ein Flußdiagramm repräsentiert werden kann, hat Fodor (1974) hat dementsprechend postuliert, dass psychologische Theorien die Form von Flußdiagrammen haben sollten.

Shagrir (2005), Abschn. 2.2.2 [Rise and Fall], hat eine interessante Interpretation der putnamschen Definition gegeben, bei der die Definition eines

[16] 1885 – 1977, Chemiker, Philosoph

[17] 1926 – 2016, Philosoph

[18] Womit nicht nur numerische Berechnungen gemeint sind, sondern auch Folgen logischer Operationen.

[19] Mentale Ereignisse können zu Typen gruppiert werden, zu denen bestimmte Typen von "physikalischer Aktivität" korrespondieren

[20] Ein Übergang $S_j \to S_k$ vom Zustand S_j zum Zustand S_k findet nicht deterministisch, also mit Wahrscheinlichkeit 1 oder 0 statt, sondern mit einer Wahrscheinlichkeit p_{jk}, die einen Wert zwischen 0 und 1 hat.

mentalen Terms als kausale Interaktion von mentalen Zuständen eliminiert wird: Shagrir gibt eine Formel an, die nur die logischen Terme 'es existiert', 'und' und Zustandbezeichnungen $S_1, \ldots, S_n$, sowie Ausdrücke für Inputs und Outputs, aber eben keine mentalen Terme enthält. Es resultiert eine Definition der *funktionalen Organisation*, die durch

$$FO(S_1, \ldots, S_n, i_1, \ldots, i_k, o_1, \ldots, o_\ell) \tag{2}$$

gegeben ist; hierin ist FO = funktionale Organisation, $S_1, \ldots, S_n$ bezeichnen die Zustände, $i_1, \ldots, i_k$ repräsentieren Inputs, und $o_1, \ldots, o_\ell$ repräsentieren Outputs. Was ein mentaler Zustand ist wird durch die funktionale Organisation festgelegt, und auf die Frage, was 'mental' genau bedeutenn soll, wird gar nicht eingegangen.

In vielen Diskussionen des Funktionalismus ist der komputationale Funktionalismus gemeint, allerdings impliziert der Funktionalismus nicht dem Komputationalismus noch ist der Komputationalismus gleichbedeutend mit Funktionalismus (Piccinini (2004)). Der Punkt bei diesen Bemerkungen ist, dass formale Definitionen wie die in (2) gegebene darüber hinwegtäuschen, dass der Begriff des Komputationalismus nur unscharf definiert ist. Die etwas kaviersmäßige Gleichsetzung von funktionalistischer Organisation mit dem Begriff 'Computer' suggeriert, dass es verschiedene Arten von Komputationalismen gibt.

Die Turingmaschine ist eine abstrakte Definition des Algorithmusbegriffs und modelliert offenbar nicht direkt das Gehirn, aber Putnam (1967) nimmt an, dass das Konzept er Turingmaschine es erlaubt bzw. ermöglicht, die Aktivität eines Gehirns als Flußdiagramm zu repräsentieren. Tatsächlich schreibt Putnam (1960), p. 159, dass er nicht postulieren wolle, dass Maschinen denken können oder dass sie eine Sprache verwenden. Andererseits besagt die Church-Turing-These, dass alles, was überhaupt "berechnet" werden kann, durch eine Turing-Maschine berechnet werden kann. Eigentlich wird also angenommen, dass kognitive Aktivitäten durch berechenbare Funktionen abgebildet werden, in diesem Sinne repräsentiert eine Turingmaschine die jeweiligen kognitiven Funktionen. Putnam folgert, dass *jede* kognitive Aktivität durch Turingmaschinen repräsentiert werden kann.

Die Zustände der Maschine (und damit des Bewußtseins, wenn mentale Ereignisse als bewußte Zustände definiert werden) nennt Putnam *logische Zustände*, sie sind sozusagen die Software der Maschine. Die physischen Zustände, die die Maschine dabei durchläuft, entsprechen der Hardware der Maschine. Eine analoge Beschreibung könne, so Putnam, für das menschliche System gefunden werden.

Für Putnam scheint sich diese Folgerung bereits aus der allgemeinen Definition einer Turingmaschine zu ergeben, denn er liefert keinerlei weitere Rechtfertigung für seine Behauptung. Aber der Turingsche Berechenbarkeitsbegriff

ist spezieller, als von Putnam angenommen wird: wie weiter oben schon angemerkt wurde, sind Berechnungen im Turingschen Sinne Folgen von Operationen aus einer diskreten Menge von Operationen, die auf Zustände aus einer diskreten Menge von Zuständen angewandt werden. Aber die Dynamik emotionaler Zustände scheint dieser Charakerisierung nicht zu entsprechen, emotionale Zustände scheinen sich eher kontinuierlich zu verändern. Und ob gedankliche Prozesse, also Folgen von Schlußfolgerungen, sauber als diskrete Folgen von Zustandsbergängen beschrieben werden können ist ebenfalls eine offene Frage. Man wird also den Begriff der Berechenbarkeit, wie er in der Logik entwickelt wurde, in Bezug auf seine Anwendbarkeit auf die Dynamik der Kognitionen diskutieren müssen (vergl. etwa Piccinini (2004), Piccinini & Shagrir (2014)). Jedenfalls ergibt sich bei der auf diskrete Zustandsmengen bezogenen Interpretation der Church-Turing-These bereits eine Art von Pankomputationalismus: alles, was eine Funktion hat, derzufolge Folgen[21] von Symbolen als Antwort auf Input-Folgen[22] erzeugt werden. Aber die Annahme, dass "alles" durch eine TM beschrieben werden kann trivialisiere die Annahme der Analogie zwischen kognitiver Dynamik und Turingmaschinen, der Pankomputationalismus erzeugt mehr Probleme als er löst: es ist ja zum Beispiel eine grundsätzliche Frage, ob es sinnvoll ist, von vorn herein anzunehmen, dass die kognitive Dynamik in diskreten Schritten verläuft, bzw. stets auf eine Beschreibung in Termen diskreter Schritte reduziert werden sollte. Ist die Dynamik kontinuierlich, so ist eine diskretisierte Beschreibung nur unter speziellen Bedingungen, etwa im Rahmen eines bestimmten Experiments, sinnvoll und dabei abhängig von der Natur der kontinuierlichen Prozesse. Shagrir (2006) zitiert Gödel: der Begriff 'Computation' sei unklarer, als es scheint, und Gödel und Turing hatten verschiedene Auffassungen:

> On the one hand, Gödel emphasizes at every opportunity that "the correct definition of mechanical computability was established beyond any doubt by Turing". On the the other hand, in a short note entitled "A philosophical error in Turing's work," Gödel [1972a] criticizes Turing's argument for this definition as resting on the dubious assumption that there is a finite number of states of mind.

Es ist bemerkenswert, dass die Möglichkeit kontinuierlicher kognitiver Prozesse in der Literatur zum Funktionalismus kaum diskutiert wird. Eine Ausnahme ist van Gelder (1998), dessen Arbeit eine ausführliche Diskussion zwischen Proponenten und Opponenten der "dynamischen Hypthese" enthält.

Putnam (1967) identifiziert den Geist (im Sinne von *mind*) mit der funktionalen Organisation im Sinne von (2) eines denkenden Organimus: die Fähigkeit, Schmerz zu empfinden setzt eine bestimmte Art von funktionaler Organisation voraus (Putnam 1967b, p. 434) Zusammenfassend läßt sich der

[21] der englische Ausdruck ist 'string', was gewöhnlich mit 'Schnur' übersetzt wird. Es macht aber im gegebenen Zusammenhang mehr Sinn, von Folgen zu sprechen.

[22] input strings

Funktionalismus wie folgt charakterisieren (Shagrir (2005): Für jede mentale Eigenschaft M existiert eine funktionale Eigenschaft F derart, dass die Aussagen

UR_F: jedes M-Ereignis ist auch ein F-Ereignis,
SUP_F: Jedes F-Ereignis ist auch ein M-Ereignis.

gelten; UR_F bedeutet einfach, dass jeder mentale Zustand auf eine dazu korrespondierende Weise funktional erzeugt wird, und umgekehrt heißt SUP_F, dass zu jedem funktionalen Ereignis ein dazu korrespondierendes mentales Ereignis existiert (Supervenienz). Ob derlei Ausführungen, ebenso wie die Arbeiten von Piccinini und Shagrir (2014) und Piccinini (2006) diskutieren den *Pan-computationalismus*, wie er von Putnam (1967)[23] eingeführt wurde: alles sei komputational, für die Neuro- bzw. Kognitionswissenschaft förderlich sind ist allerdings fraglich: es geht um Explikationen von Begrifflichkeiten, auf die in der konkreten Forschung so gut wie nie Bezug genommem wird.

Hinweis auf Lucas-Frey: Abschnitt 2.4.2

David Marrs Komputationalismus: Marr & Hildreth (1980) und Marr (1982) haben eine andere Art von Komputationalismus eingeführt, die sich von der Putnams grundsätzlich unterscheidet und nicht nur für Fragen zur Wahrnehmung von Interesse ist. Sie bezieht sich auf die Wahrnehmung bzw. Entdeckung visueller Reize. Die Frage, ob die Wahrnehmung der Stimuli eine neuronale Basis hat, wird gar nicht erst gestellt, es wird vielmehr das neuronale Aktivierungsmuster modelliert.

Ein visuelles Muster oder eine visuelle Szene wird u.a. durch eine Funktion definiert, die Helligkeitsveränderungen beschreibt. Auf diese Veränderungen reagieren neuronale Strukturen mit unterschiedlichen Skalen, die den Verlauf der Veränderungen abbilden. Die erste Ableitung dieser Funktion beschreibt die Veränderung etwa von Helligkeit, und wo die Veränderung – also die erste Ableitung – gleich Null ist, nimmt die Funktion ein Extremum an, und die Veränderung der Veränderung, also die zweite Ableitung, kreuzt die x-Achse, wobei x die Helligkeit repräsentiert. Sind x und y retinale Koordinaten, so ist die zweite Ableitung durch die Laplacesche der Funktion gegeben[24]

$$\nabla^2 = \delta^2/\delta x^2 + \delta^2/\delta y^2 \tag{3}$$

Marr nimmt also an, dass die Wahrnehmungsprozesse mathematisch beschrieben werden können und dementsprechend wohldefinierten Gesetzmäßigkeiten unterliegen, und er diskutiert diese Annahme auch nicht weiter. In der Philosophie des Geistes wird dagegen diese Annahme oft als fragwürdig angesehen.

[23] Nature of mental states

[24] Einen guten Überblick über den Marrschen Ansatz findet man in https://en.wikipedia.org/wiki/David Marr

So argumentierte Davidson (1970), dass es überhaupt eine psychophysischen Gesetze gebe. Er konstatiert, dass mentale Ereignisse wie Wahrnehmungen, Erinnerungen, Entscheidungsprozesse und Handlungen "nicht im nomologischen Netz einer physikalishen Theorie eingefangen" werden können. Andererseits folgt er der Putnamschen Annahme, dass zwischen mentalen Ereignissen kausale Beziehungen bestehen, woraus folge, dass mentale Ereignisse eine Anomalie darstellen, – kausale Determiniertheit und Anomalität seien unbezweifelbare Fakten, wobei er sich auf Kants Theorie der Freiheit mentaler Akte bezieht. Dieser Widerspruch sei aber nur scheinbar: er postuliert drei Prinzipien: (i) zumindest einige mentale Ereigniss interagieren mit physischen Ereignissen: wenn ein Seemann sieht, dass sich das Schlachtschiff *Bismarck* nähere, so liegt das daran, dass sich das Schlachtschiff *Bismarck* nähere. (ii) Dem zweiten Prinzip zufolge impliziere, dass Kausaltität die Existenz deterministischer Gesetzmäßigkeiten impliziert. (iii) das dritte Prinzip – das Prinzip der Anomalität des Mentalen – besagt, dass es keine strikten, deterministischen Gesetze gebe, die mentale Ereignisse vorherzusagen und zu erklären gestatten. Dabei will er keinen reinen Dualismus propagieren: mentale Ereignisse würden auf physische Ereignisse *supervenieren*, d.h. es können keine zwei Ereignisse existieren, die in allen physischen Aspekten übereinstimmen, sich aber hinsichtlich der korrespondierenden mentalen Ereignisse zumindest in einigen Aspekten voneinander unerscheiden (p. 141). Im Kapitel über den freien Willen wird auf dieses quasi-dualistische Argument näher eingegangen.

Auf derlei philosophische Ansätze geht Marr also gar nicht ein; die Frage, ob psychophysische Gesetze überhaupt möglich sind, wird also gar nicht erst gestellt, – aber Marr ist ja auch kein Philosoph. Das ist aber nur ein Punkt. Der zweite ist, dass der Zustandsbegriff von dem im Putnamschen Ansatz definierten ein anderer ist. Während beim Putnamschen Ansatz die Zustände eine diskrete, also abzählbare Menge bilden, ist diese Menge beim Marrschen Ansatz überabzählbar, – weil kontinuierliche Trajektorien betrachtet werden. Der Grund dafür liegt daran, dass bei Marr die Zustandsmenge durch ein dynamisches System definiert ist, dass durch im Prinzip durch eine Differentialgleichung beschrieben werden kann, wobei oft die Koeffizienten in guter Näherung Konstante sind.

Dynamische Systeme werden als Alternative zu dem von Putnam postulierten Ansatz gesehen, der wiederum oft in Zusammenhang mit dem Prinzip der Multiplen Realisierbeit gebracht wird; Turingmaschinen repräsentieren ein Prinzip, das nicht an einen bestimmten physikalischen Rahmen gebunden ist. Dynamische Systeme werden einerseits – analog zur Turingmashine – abstrakt durch allgemeine mathematische Prinzipien definiert, können andererseits sehr spezielle reale Systeme beschreiben, d.h. sie können direkt für die Beschreibung neuronaler (Teil-)Systeme spezifiziert werden (Dayan & Abbott (2001), Gerstner & Kistler (2002)).Ebenso ist es möglich, dynamische Systeme auch für allgemeine psychologische Zustände und ihre Interaktion zu formulieren.

Bevor darauf eingegangen wird soll noch ein für das Modell der funktionalen Organisation als besonders wichtig angesehenes Merkmal vorgestellt werden, die den Funktionalismus für viele Philosophen erst interessant gemacht hat: das der Multiplen Realisation.

2.3 Funktionale Organisation und Reduktion

2.3.1 Multiple Realisierbarkeit

Für Philosophen scheint dann auch der Reiz der funktionalistischen Definition mentaler Zustände insbesondere darin zu bestehen, dass im Gegensatz zu den Gehirn-Geist-Identitätstheorien von Place und Smart eben gerade kein expliziter Bezug auf eine neurale Basis genommen wird: Tiere, vom Schimpansen über Hund und Katze bis hin zum Wurm und zur Molluske hätten schließlich ganz unterschiedliche Hirnstrukturen, könnten aber gleichwohl Schmerz – der ja ein mentaler Zustand ist – empfinden. Ein mentales Ereignis wie die Empfindung von Schmerz sei also auf verschiedene Weise erzeugbar, und schließlich sei es denkbar, dass Aliens ("extraterrestrische, mit grünem Schleim durchpulste Wesen") mit einer "Hardware" versehen seien, die sich von unseren Gehirnen in grundsätzlicher Weise unterscheiden könne. Putnams Version des (komputationalen) Funktionalismus ist speziell am Modell der Turingmaschine konzipiert worden, und ein wichtiger Aspekt der Turingmaschine ist, dass sie nicht in Bezug auf eine spezielle physische Realisierung gedacht werden muß. Das Wichtige an der Turingmaschine ist, dass ihre Funktion bzw. ihr Verhalten durch ein Programm definiert ist, und das Verhalten wird bestimmt durch
1. den Input zu einer gegebenen Zeit,
2. den Zustand der Maschine zu dieser Zeit,
3. die Zustandstabelle.
Die Zustandstabelle enthält in einer gegebenen Zeile (a) den Zustand S_t zur Zeit t, (b) den Input, (c) den Output, (d) den Zustand S_{t+1} im Zeitpunkt $t+1$. Die Tabelle repräsentiert das Programm der Maschine. Der physische Bau der Maschine ist aber nicht notwendig, was zählt, ist die abstrakte Beschreibung, und es ist dieser Aspekt, auf den sich der Funktionalismus bezieht. Er ist eine "high-level"-Beschreibung kognitiver Prozesse, der Bezug auf eine spezielle physische Realisierung ist nicht nötig. Dieser Sachverhalt ist der Hintergrund der von Putnam (1967) eingeführten These der *multiplen Realisierbarkeit*, d.h. psychologische Prozesse können im Rahmen des Funktionalismus ohne Bezug auf konkrete Hirnprozesse beschrieben werden. Dieser Aspekt wird weiter unten ausführlich diskutiert werden.

Wie Bechtel & Mundale (1999), p. 76, anmerken, sind die Argumente für die MR eher intuitiv, d.h. sie bleiben im Bereich des bestenfalls Hypothetischen, ohne den Versuch, Alternativen auszuschalten oder doch zumindest

als wenig wahrscheinlich erscheinen zu lassen. Sie basieren neben dem Bezug auf die Repräsentation kognitiver Aktivitäten durch Turingmaschinen auf der Beobachtung, dass im Rahmen der aufkommenden Forschung zur Künstlichen Intelligenz (KI) Computerprogramme geschrieben wurden, die kognitive Operationen analog zu denen von Menschen durchführen. Das bedeutet, dass die Kognitionen in verschiedenen Realisierungen möglich sind. Insbesondere Philosophen gelangten zur Überzeugung, dass diese kognitiven Aktivitäten auch in Gehirnen von Außerirdischen mit gänzlich anders realisierten Gehirnen durchgeführt werden können. Am Ende kam man zu der metaphysischen Forderung, dass mentale Operationen überhaupt nicht mit biologischen oder was für Substanzen auch immer zusammengebracht werden sollten.

Dem naturalistischen Ansatz in der Kognitionswissenschaft zufolge werden mentale Zustände durch das Gehirn erzeugt. In den Identätstheorien wird speziell postuliert, dass jedem mentalen Zustand ein dazu korrespondierender Zustand des Gehirns entspricht ("brainstate theory"). Die Identitätstheorie impliziert, dass ein physikalistisches (biochemisches) Substrat[25] existiert, das in all den verschiedenen Hirnstrukturen von der Molluske bis zum *homo sapiens* vorkommt, – man muß hier bedenken, dass die Identitätstheorien in den fünfziger Jahren des zwanzigsten Jahrhunderts entstanden sind, seit dem sind viele hirnanatomische und -physiologische Kenntnisse hinzu gekommen. Putnam argumentiert, dass der von ihm konzipierte Funktionalismus derlei Rückführung auf das Physische (die Neurobiologie, die Neurophysiologie, die Biochemie) nicht zulasse, denn jedes mentale Ereignis superveniere auf ein funktionales Ereignis. Dabei müsse man auch die Möglichkeit betrachten, dass außer Menschen auch Affen, Kröten und Würmer z.B. Schmerz empfinden können, und diese Möglichkeit womöglich auch auf siliziumbasierte Androide und "extraterrestrische, mit grünem Schleim durchpulste Wesen" (Bickle 2020) zuträfe, so dass es plausibler sei, verschiedene physikalisch-chemische Substrate oder aber gleich gar keine derartigen Substrate anzunehmen.

Putnam und Fodor haben den Funktionalismus zwar als "physikalistische" Theorie interpretiert, d.h. sie nahmen an, dass mentale Ereignisse letztlich eine naturalistische Basis haben, aber Putnam fügte an, dass der Funktionalismus prinzipiell auch mit dem Dualismus verträglich sei. Wenn also der Putnamsche Funktionalismus die Rückführung auf das Physische tatsächlich nicht zuläßt, so folgt aber noch lange nicht, dass mentale Ereignisse tatsächlich nicht rückführbar sind. Man kann ebensogut die Hypothese der Rückführbarkeit als Arbeitshypothese akzeptieren und damit den Funktionalismus verwerfen, zumal der funktionalistische Ansatz im Licht neuerer Daten sowieso nicht attraktiv ist.

Für die Diskussion der MR-These werde noch ein Blick auf die formalisierte Fassung dieses These geworfen. Sie wird transparenter, wenn sie in der von

[25] "physical-chemical kind"

Shagrir (1998) formalisierten Form präsentiert wird. Das Zeichen $\forall$ stehe für 'für alle", Mx stehe für "x hat das Merkmal M und $M_1x \rightarrow M_2x$ stehe für "Wenn x das Merkmal M_1 hat, dann hat x auch das Merkmal M_2. So kann ein psychologisches Gesetz in der Form

$$\forall x(M_1x \rightarrow M_2x) \tag{4}$$

geschrieben werden: "Für alle x gilt, wenn M_1x gilt, dann auch M_2x. M_1 und M_2 sind psychologische Prädikate ("Typenprädikate"). Weiter sei ein physikalisches Gesetz der Form

$$\forall x(P_1x \rightarrow P_2x) \tag{5}$$

gegeben, wobei P_1, P_2 physikalische (Typen–)Prädikate sind. Die Frage ist nun, ob zu Gesetzen der Form (4) auch Gesetze der Form (5) korrespondieren. Für eine derartige Korrespondenz muß ein *Brückengesetz* der Art

$$\forall x(M_ix \rightarrow P_ix) \tag{6}$$

gelten. Das MR-Prinzip besagt nun, dass einem psychologischen Prädikat M verschiedene Prädikate $P_1, P_2, \ldots$ der physikalischen Art entsprechen, was in der Form

$$\forall x(Mx \leftrightarrow P_1x \vee P_2x \vee P_3x \vee \ldots) \tag{7}$$

ausgedrückt werden kann. Das Zeichen $\leftrightarrow$ steht für Bikondionalität ("dann und nur dann"). Der Ausdruck rechts des Zeichens $\leftrightarrow$ ist eine von Fodor (1974) so genannte "wilde" Disjunktion (eine Disjunktion ist eine Oder-Verbindung, das Zeichen $\vee$ steht für lat. *vel*, womit das einschließende Oder gemeint ist) von physikalischen Prädikaten: x hat das Prädikat P_1 oder x hat das Prädikat P_2, etc. "Wild" heißt dabei, dass es kein physikalisches Prädikat gibt, dass allen gemein ist, und dass es im Prinzip unendlich viele P_i gibt. (7) entspricht dem umgangssprachlich formulierten Prinzip der Multirealisierbarkeit, verdeutlicht aber eine in der umgangssprachlichen Formulierung nicht gleich aufscheinende Implikation. Die Disjunktion $P_1x \vee P_2x \vee P_3x \vee \ldots$ soll aussagen, dass x das mentale Ereignis M hat, wenn x eines der "physikalischen" Prädikate P_i hat. Gleichzeitig bedeutet die Oder-Verbindung $\vee$ in der gewöhnlichen Logik, dass es egal ist, welches der Prädikate P_i auf x zutrifft, und da die Anzahl der möglichen Prädikate nicht begrenzt ist – es soll ja auch der von grünem Schleim durchpulste Astronaut von Alpha Centauri das mentale Ereignis M haben können – macht man keine Annahmen über die Anzahl der P_i.

Die Idee der Multiplen Realisierbarkeit wurde bald zur "neuen Orthodoxie" der Philosophie des Geistes, zusammen mit der Lehre des *nichtreduktiven Physikalismus*[26]), der davon ausgeht, dass mentale Ereignisse einerseits nicht in einer immateriellen Substanz realisiert werden, aber andererseits auch

[26]Physikalismus meint einen Materialismus, der nicht mit dem Atheismus assoziiert wird, es ist nur gemeint, dass Ereignisse irgendwie physikalisch sind.

nicht auf physikalische Prozesse reduziert werden können (Godbey (1978). Die Spezifikation der multiplen Realisierbarkeit in (7) durch eine Disjunktion ist allerdings nicht so klar, wie ihre formale Struktur suggeriert. In der Logik wird eine Disjunktion genau dann als "wahr" definiert, wenn mindestens eine der Alternativen P_j "wahr" ist. Gleichzeitig soll (7) als Argument für die Nichtreduzierbarkeit des Mentalen dienen. Vielleicht ist dies ein Grund für die Bezeichnung 'Wilde Disjunktion': gemeint ist nicht die übliche Definition, sondern eine Definition, die mit dem nichtreduktivne Physikalismus kompatibel ist. Das Konzept des nichtreduktiven Physikalismus ist wohl, wie das der Anomalität des Mentalen, das Resultat des Versuchs, mit der Schwierigkeit, dass Mentale explizit auf neuronale Prozesse zurückführen zu können, zurecht zu kommen. Vom Standpunkt der empirischen Wissenschaft aus gesehen wird man sich empirisch bemühen müssen, die Rückführbarkeit kognitiver Funktionen auf biologische Prozessse nachzuweise, oder zu widerlegen. Philosophen sehen ihre Aufgabe anscheinend darin, die Adäquatheit dieser These deduktiv aus allgemeinen Begrifflichkeiten herzuleiten. So wird postuliert, dass keine der Alternativen zutreffen könne, so dass man schließen müsse, dass mentale Ereignisse nicht physisch basiert seien[27]. Sicherlich kann man die Identitätstheorie kritisieren, wenn man sie so interpretiert, dass sie die *Identität* von mentalen und "physikalischen" (neurophysiologischen) Prozessen fordert, z.B. sei Schmerz *identisch* mit der Aktivität von C-Fasern, – aber genau soll hier 'identisch' bedeuten? Blockiert man die C-Fasern, so wird möglicherweise kein Schmerz mehr empfunden, aber das heißt doch nur, dass das Feuern von C-Fasern eine notwendige, aber deswegen doch nicht auch hinreichende Bedingung für die Schmerzempfindung ist. So schreibt zum Beispiel Beckermann (2008, p. 137), die MR-These habe der Identitätstheorie "den empirischen Boden unter den Füßen weggezogen": bei Positronen-Emmissions-Tomographien, so Beckermann, ergäben sich zwar ähnliche, aber eben nicht identische Muster, und außerdem gebe es statistisch signifikante Unterschiede zwischen Männern und Frauen. Auch bei verschiedenen Untersuchungen an ein und derselben Person würden die Korrelationen zwischen mentalen Ereignissen und neurophysiologische Mustern stark variieren, und darüber hinaus sei die wohldokumentierte Plastizität der Hirnfunktionen ein Argument gegen die Identitätstheorie. Nein, diese Sachverhalte zeigen nur, dass die Identitätstheorien, auf die Beckermann sich bezieht, inadäquat oder einfach unvollständig formuliert wurden. Weiter stellt Beckermann die rhetorische Frage, ob Marsmenschen oder Roboter schon deswegen kein mit unserem mentalen Leben vergleichbares Leben haben könnten, weil ihre Hirne siliziumbasiert und nicht biologisch seien. Dieses Argument wird oft mit dem Begriff des Chauvinismus assoziiert: weil Tiere Hirnstrukturen haben, die sich von den Strukturen unserer Gehirne unterscheiden und Aliens ebenfalls ganz andere Hirnstrukturen haben könnten sei es chauvinistisch, mentale Ereignisse mit unseren Hirnstruktu-

[27] vergl. Kim (1972) für eine vertiefende Diskussion.

ren zu identifizieren. Diese Betrachtungen machen es, so Beckermann, "mehr als unwahrscheinlich, dass mentale und physische Zustände" die Bedingungen erfüllen, die Beckermann (p. 109) so formuliert hat:

Identität: (*) F und G sind identisch, wenn für alle x gilt: x hat F genau dann, wenn x G hat, d.h. $\forall x(Fx \leftrightarrow Gx)$, woraus sich die notwendigen Bedingungen für Identität ergeben:
(i) F und G sind *koextensional*, d.h. (*) ist wahr,
(ii) F und G sind *nomologisch koextensional*, d.h. (*) ist naturgesetzlich wahr.

Die Frage ist, was mit solchen Definitionen erreicht werden soll, außer, dass bestimmte Thesen/Hypothesen "widerlegt" werden können. Die Frage, was Identität bedeuten soll, ist berechtigt - schon Leibniz hat sie gestellt – aber die Frage ist doch, ob der Identitätsbegriff in der hier gegebenen Fassung wirklich relevant ist, wenn es um die Rückführbarkeit psychischer auf neurophysiologische Prozesse geht. Die formale Strenge einer Definition kann nützlich sein, kann aber auch die philosophische Argumentation über Gebühr verengen: letzlich geht es bei den Betrachtungen um die Struktur natürlicher Prozesse, die sich i. A. nicht rein deduktiv aus philosophisch konzipierten Grundannahmen erschliessen lässt: dies ist ja einer der Gründe, weshalb sich die empirische Wissenschaften entwickelt haben. Mentale und neurophysiologische Zustände mögen in der Tat nicht miteinander identisch sein, und die von Place, Smart und Feigl formulierte Identitätstheorie mag in ihrer in den 50er Jahren des 20ten Jahrhunderts entstandenen Version nicht zutreffend sein. Dafür gibt es neuere Ansätze, etwa der von Singer (2007), die auf dem empirischen Befund beruhen, dass kognitive Akivitäten auf räumlich verteilte und parallel arbeitende Neuronenverbände zurückzuführen sind; die Beschreibung derartiger Neuronenpopulationen unterscheidet sich wesentlich von den Ansätzen von Smart und Place.

2.3.2 Reduzierbarkeit von Theorien

Fodor hat in seiner Arbeit *Special Sciences* (1974) argumentiert, dass die multiple Realisierbarkeit mentaler Zustände die Nichtreduzierbarkeit der "high-level" Wissenschaft Psychologie auf die Neurophysiologie impliziert. Es mache keinen Sinn, neurophysiologische Mechanismen für psychologische Prozesse suchen zu wollen, die Psychologie bzw. die Kognitionswissenschaften seien *methodologisch autonom*. Sicher könne man Untersuchungen über die Beziehungen zwischen mentalen und neurophysiologischen Zuständen anstellen, aber sie seien letzlich uninformativ und daher überflüssig. Beckermann zitiert Fodor (1974) mit der Aussage, dass bestimmte Begriffe ("Artbegriffe") aus Einzelwissenschaften – Geologie, Psychologie, Soziologie – nicht in eindeutiger Weise bestimmten Artbegriffen der Physik entsprechen, woraus folge,

dass diese Wissenschaften nicht auf die Physik reduziert werden können; das ausführliche Zitat kann in Beckermann (2006, p. 139) gefunden werden. Es sei klar, so Beckermann,

> "dass die Identitätstheorie auf völlig hoffnungslosem Posten steht, wenn Fodor mit seinen Überlegungen Recht hat. Und alles spricht dafür, dass er damit Recht hat. Die Multirealisierbarkeit mentaler Zustände und Eigenschaften gilt daher heute als das entscheidende Argument gegen die Identitätstheorie". (Beckermann, p. 141).

Diese These hat viele Anhänger gefunden. Fodors Argumentation macht allerdings von einer besonderen, auf eine Arbeit Nagels (1961) zurückgehende Definition von Reduzierbarkeit Gebrauch, die lange Zeit als Standarddefinition galt. Sie wird im Rahmen der Kritik an der These der multiplen Reduzierbarkeit und von Fodor postulierten Nichtreduzierbarkt in Abschnitt 2.4.3 vorgestellt.

Bevor auf die Problematik der Reduzierbarkeit[28] näher eingegangen wird, soll noch ein Argument *gegen* den Funktionalismus aus Putnams eigener Hand vorgestellt werden, womit der Übergang zur Kritik des komputationalen Funktionalismus vorbereitet wird.

2.3.3 Putnams Theorem

Putnam wurde mit den Jahren seinem Ansatz, dem Funktionalismus, gegenüber immer skeptischer und bewies schließlich (1988) ein Theorem:

Putnams Theorem Jedes gewöhnliche offene System ist die Realisierung von jedem abstrakten, endlichen Automaten.[29]

Gilt dieses Theorem, so hat auch ein Felsbrocken Geist (im Sinne von 'mind'), vorausgesetzt, der Geist funktioniert nach Maßgabe des Funktionalismus. Nun sind wir uns sicher, dass ein Felsbrocken keinen kognitiven Prozessor hat, so dass wir schließen können, dass der Funktionalismus keine adäquate Theorie ist. Man kann sagen, dass alles, was den Automaten realisiert, ein kognitives System hat. Sollte also das Theorem wahr sein, so kann mein Gehirn unendlich viele verschiedene funktionale Organisationen implementieren, jede mit einem anderen kognitiven System (Shagrir (2005)).

Damit unterminiert das Theorem das in Abschnitt 2.2.2 vorgestellte Prinzip SUP_F, einmal, weil dem Theorem zufolge ein Felsen mein kognitives System implementieren kann, aber sicher kein kognitives System hat, und zum

[28] Beckermanns Urteil ist voreilig (s. oben, letzte Bemerkung); bereits 1999 haben Bechtel und Mundale eine ausführliche Diskussion der Beziehung zwischen mentalen und neuronalen Zuständen vorgelegt.

[29] In: Representation and Reality. Cambridge (Mass.), London 1988, pp. 121 – 125

anderen, weil mein Gehirn dem Theorem zufolge viele funktionale Organisationen implementiert, und jede davon konstituiert ein von den anderen unabhängiges kognitives System, – aber ich habe nur eines! Diesen Befund nennt Shagrir das *Realisierungsproblem.*

2.4 Kritik und Kommentare

2.4.1 Funktionalismus allgemein

Außer dem Hinweis, dass gemäß der Church-Turing-These alle algorithmisch darstellbaren kognitiven Prozesse durch eine Turingmaschine abgebildet werden können gibt es bei Putnam und Fodor keine Herleitung oder tiefere Begründung dieser These. Einige Annahmen des Funktionalismus sind aber nicht so selbstverständlich, wie Putnams Darstellung es suggeriert:

1. Funktionale Zustände bilden eine *diskrete Menge*,
2. Art der Definition des Begriffs des mentalen Ereignisses als Resultat einer funktionalen Interaktion,
3. Die funktionale Organisation FO ist kausal,

Polger[30] merkt an, dass gemäß Putnams Definition ein mentaler Zustand eher durch das, "was der Zustand macht, als woraus er besteht"[31] charakterisiert wird, sein "Wesen" besteht darin, andere Zustände zu erzeugen, die wiederum andere Zustände erzeugen, etc. Die Zustände erklären sich gegenseitig, sie bilden sozusagen ein sich selbst tragendes System.

Nun ist ein wichtiger, aber problematischer Aspekt des funktionalistischen Ansatzes die in Punkt 1. postulierte *diskrete* Menge von Zuständen, auf die im Folgenden immer wieder zurückgekommen wird. Zunächst stellt sich die Frage, in welcher Weise die Annahme diskreter Zustandsmengen zum Beispiel mit der phänomenologisch erlebten kontinuierlichen Variation mentaler Zsutände vereinbar ist. Ein erstes, einfaches Beispiel ist die Intensität von Sinneseindrücken. Bereits Fechner[32] argumentierte 1860, dass die Empfindungsstärke logarithmisch und damit *stetig* mit der Stimulusintensität wächst (diese Funktion ist die *psychophysische Funktion*). Die Aussage, dass die Empfindungsstärke *logrithmisch* von der Stimulusintensität abhängt, mag bezweifelt werden, es ist eher die Stetigkeit psychophysischen Funktion, die hier wichtig ist: angenomen, sie sei *nicht* stetig. Dann enthält die Funktion "Sprünge", d.h. es existieren Veränderungen der Funktion, die ohne zeitliche Dauer stattfinden[33]. Diese Eigenschaft ist biologisch unplausibel. Auch die von Stevens[34] als Alternati-

[30] Thomas W. Polger "Functionalism", *Internet Encyclopdiea of Philosophy (IEP)*

[31] "... what makes something a mental state is more a matter of what it does, not what it is made of".

[32] Gustav Fechner (1801 – 1887)

[33] Zum Beispiel: $g(t) = f(t)$, $t \leq t_0$, $g(t) = h(t)$, $t > t_0$, $h(t) - f(t) \neq 0$ für $t = t_0$,

[34] Stanley S. Stevens (1906 – 1973)

ve zur Fechnerschen Funktion vorgeschlagene Potenzfunktion postuliert die Stetigkeit der Empfindungsstärke: es ist schwer, eine plausible Erklärung für Unstetigkeiten zu finden. Putnam geht auf derlei Fragen gar nicht ein; akzeptiert man seine Zustandsdefinition, so würde man annehmen, dass die Empfindungsskala durch eine Treppenfunktion gegeben ist: auf jeder Stufe bleibt die Empfindung trotz steigender Stimulusintensität konstant und die Veränderung der Funktion von Stufe zu Stufe ist unstetig im mathematischen Sinn. Tatsächlich könnte man so argumentieren, wenn man das Webersche Gesetz[35] so interpretiert, dass für Intensitäten ψ innerhalb des Intervalls $(x, x + \Delta x]$ (also $x < \psi \leq x + \Delta x$) die Empfindungsstärke nicht variiert. Nur widerspricht diese Annahme dem empirischen Befund, dass Wahrnehmungsschwellen keine physiologische Konstanten, sondern kontinuierliche Größen sind. Für viele Philosophen des Geistes stellt dieser Befund aber schon deswegen kein Problem dar, weil sie sich auf Donald Davidsons (1970) Argument, dass es gar keine psychophysischen Gesetze (im strengen Sinn) geben könne berufen, Davidson zufolge gleicht die Fechner-Debatte einer Spiegelfechterei. Analoge Aussagen lassen sich über emotionale Intensitäten machen: Emotionen können sich sicher sehr schnell ändern, aber mit großer Wahrscheinlichkeit nicht unstetig. Ob die Formierung von Gedanken im Rahmen von diskreten Schritten erfolgt, wie sie zum Beispiel durch die schriftliche Fassung eines Lösungsweges für ein Problem oder eines mathematischen Beweise suggeriert wird oder ob sie einer kontinuierlichen, leider nicht direkt beobachtbaren Dynamik folgt, ist eine introspektiv schwer zu entscheidende Frage, – bestimmte Gedanken kommen, oder sie kommen nicht, und introspektiv hat man i. A. keinen Zugang zu den Prozessen, die das Kommen oder Nichtkommen eines Gedanken bewußt werden lassen. Geht man davon aus, dass der Lösungsprozess durchgehend ein neuronaler Prozesse ist, so ist es schwieriger, diesen als einen prinzipiell diskreten Prozess zu beschreiben, – wie soll eine neuronales Systen in strengem Sinne unstetige Sprünge erzeugen? Es liegt also nahe, kognitive Prozesse im Rahmen der Theorie kontinuierlicher dynamischer Systeme zu modellieren bzw. zu verstehen. Im Kino wird eine Folge von minimal 18 Bildern pro Sekunde als kontinuierliche Bewegung gesehen, wären mentale Zustände grundsätzlich diskret, so würden wir die Welt grundsätzlich nur verruckelt sehen, – was offenbar nicht der Fall ist.

In Abschnitt 2.1 wurde auf die rekursive Struktur des Funktionalismus hingewiesen: ein mentaler Zustand entsteht durch Interaktion von mentalen Zuständen, die ihrerseits als Interaktion von mentalen Zuständen definiert sind. Wie sind denn die Zustände definiert, deren Interaktion die mentalen Zustände erzeugt? Eben durch die Interaktion von noch basaleren Zuständen, und so weiter ... – um einen infiniten Regress zu vermeiden, könnte man die Existenz von Basisemotionen postulieren, deren Kombinationen die Menge der möglichen Emotionen definiert. Diese Annahme wird durch empirische Befun-

[35] das nicht in der von Fechner angenommenen Form gilt

de gestützt: Furcht, Freude, Trauer, Ekel und Ärger werden als solche Basisemotionen diskutiert[36]. Soll die funktionalistische Definition allgemein gelten, so muß sie auch für die Basisemotionen gelten: z.B wäre dann der Zustand der "Furcht" durch eine spezielle Konstellation dieser Basisemotionen definiert, und man kommt auch nicht umhin, zu erklären, was denn eine Basisemotion ist, sollte sie keine funktionalistische Kombination sein. Sollten Basisemotionen sozusagen außerfunktionalistisch definiert sein, so wäre der Funktionalismus eine unvollständige Theorie, die durch eine nicht-funktionalistische Annahme erweitert werden müsste und damit das funktionalistische Konzept selbst in Frage stellen würde. Plausibler wäre ein Ansatz, dem zufolge sich Basiszustände direkt in bestimmten Sitationen ergeben und so den Gesamtzustand als Mischung von Basiszuständen darstellen, wobei man noch die zusätzliche Annahme machen könnte, dass die Basiszustände sich unabhängig voneinander einstellen können, – aber das sind Details, die hier nicht diskutiert werden müssen. Nun geht es bei Kognitionen aber nicht nur um emotionale Zustände, sondern auch um Aktivitäten, also zum Beispiel um Denkvorgänge. Man kann sagen, dass eine Person in einem bestimmten Zustand ist, wenn sie eine Schlußfolgerung vollzieht. Oder dass das Ziehen einer Schlußfolgerung eine kontinuierliche Folge von Zuständen ist. Diese Annahme ergibt sich aus der Annahme, dass Denkprozesse durch eine kontinuierliche Dynamik charakterisiert sind. Bei einem Blick aus dem Fenster nimmt eine Person etwa wahr, dass die Strasse trocken ist und folgert, dass es nicht regnet. Sie könnte auch eine Reihe von anderen Schlüssen zieht, – welche sie zieht, wird vom kognitiven Gesamtzustand abhängen. Im Rahmen eines üblichen Forschungsprogramms wird man nun versuchen, z.B. Mechanismen der Aufmerksamkeitsfokussierung und mglicherweise noch andere Prozesse, die die jeweiligen Schlußfolgerungen mit bestimmen, direkt zu untersuchen. Dabei wird wird die Problematik des Putnamschen Funktionalismmus deutlich: wenn die Endlichkeit der Rekursion auf die implizit postulierte Existenz von mentalen Basiszuständen führt, so läßt sich kaum vermeiden, ihre Entstehung erklären zu müssen. So erklärt zum Beispiel der Hinweis auf die Aktivierung farbspezifischer Neurone in Area V4 des visuellen Kortex' noch nicht, wie auf der Basis dieser Aktivierung die Wahrnehmung "rot" entsteht. Im Gehirn werden ständig Teilpopulationen von Neuronen aktiviert, ohne dass die erzeugten Spike-Folgen ein dazu korrespondierendes bewußtes Erleben bedeuten, und analoge Anmerkungen gelten für das Erleben und Erzeugen von Gedanken, von Emotionen, etc. Erzeugt ein visueller Stimulus etwa die neuronale Aktivierung N_k, und ist S_r die Wahrnehmung 'rot', so wird die Bedeutung der Aussage $N_k \rightarrow S_r$ durch die Interaktion mit anderen Zuordungen von neuronalen Aktivitäten und Wahrnehmungen erklärt, und es stellt sich die Frage, wie diese Weckselwirkungen beschrieben werden können. Generell wird nicht erklärt, welcher Erkenntnisgewinn sich aus Definitionen der Art (2), Seite 16, ergibt. Die Definition des

[36] vergl. Spektrum - Lexikon der Neurowissenschaft

Funktionalismus sagt nur, dass es Wechselwirkungen gibt, aber nicht, wie sie jeweils funktionieren, und dass es eine Dynamik von Zustandsübergängen gibt wurde von Psychologen lange vor der Konzeptualisierung des Funktionalismus angenommen.

Die Beschreibung einer Dynamik gemäß (2), Seite 16, mit einer Beschränkung auf eine explizit diskrete Menge von Zuständen macht allenfalls als Appproximation an die tatsächliche Dynamik einen Sinn, und zwar wenn man Experimente durchführt, bei denen die neuronale Dynamik nicht direkt beobachtet werden kann. Ein Beispiel sind Experimente über Gedächtnisprozesse, bei denen in einem Versuchsdurchgang nur festgestellt werden kann, ob ein "Item", also etwa ein von der Vp zu memorisierendes Wort, oder eine Silbe oder auch nur ein Buchstabe bzw. Symbol während des Beobachtungsintervalls $(t, t+\Delta t)$ erinnert wurde oder nicht. Selbst wenn dann die unterliegenden Prozesse in Bezug auf die Daten als diskrete Prozesse – beliebt sind Markov-Prozesse – modelliert werden, so bleibt doch klar, dass man es mit Approximationen an die "wahren" Prozesse zu tun hat. Aber bei (2) ist von Approximationen nicht die Rede. (2) ist eine metaphysische Annahme über die funktionale Organisation selbst. Ein einfaches Beispiel ist die zuerst von Dollard und Miller (1939) vorgebrachte (Hypo-)These, dass eine Aggression stets eine Folge vorangegangener Frustration sei und die Intensität der erlebten Aggression proportional zur Intensität der erlebten Frustration sei. Funktionalistisch ausgedrückt besagt die These, dass Aggression kausal durch Frustration erzeugt wird. Dass die Dollard-Millersche (Hypo-)These empirische Tests nicht überlebt hat – Frustrationen erzeugen nicht immer Aggressionen, und wenn Aggressionen erzeugt werden, ist ihre Intensität nicht notwendig proportional zur Intensität der vorausgehenden Frustration, d.h. es gibt keine fixe Proportionalitätskonstante, uns bereits die für eine empirische Überprüfung der These notwendige Messung der Intensitäten von Emotionen ist problematisch. Die Dollard et al - Hypothese ist schlicht zu simpel, weil jede Vernetzung zum Beispiel mit anderen Emotionen fehlt, die die Modulation der Beziehung zwischen Frustration und Aggression in Rechnung stellen könnte. Allein dieser Sachverhalt stellt die Behauptung der Proportionalität der Intensität von Emotionen in Frage. Charakterisierungen wie (1) oder (2) sind offenbar begrifflich viel zu grobkörnig, um derartige Modulationen abbilden zu können[37].

Man könnte argumentieren, dass der Begriff der funktionalen Organisation doch nützlich sei, weil damit die Möglichkeit gegeben werde, die jeweilige Dynamik der Beziehung zwischen Frustration und Aggression in ihrer Einbettung in die Gesamtdynamik von Emotionen zu klären, – diese Gesamtdynamik würde erklären, dass manchmal eine Aggression als Folge einer vorausgegangenen Frustration beobachtet werde, und manchmal eben nicht. Aber dass es eine solche Einbettung gibt, wissen nicht nur empirisch arbeitende Psychologen längst, diese Einsicht ist bereits Bestandteil der Alltagspsychologie. Deshalb

[37]Eine grundsätzliche Darstellung findet man in Wissenschaftstheorie IV (1), Kapitel 1.

trägt der Begriff der funktionalen Organisation, für sich genommen, nichts zur konkreten Forschung bei: es geht darum, die funktionale Organisation explizit zu machen. Tatsächlich müssen ganz andere Fragen beantwortet werden, zum Beispiel die nach der Messbarkeit der Intensitäten von Emotionen, oder welche Rolle stochastische Fluktuationen im Gesamtsystem der Kognitionen spielen, etc. Wenn Definitionen wie (2) (Seite 16) als Teil eines philosophischen Rahmens (entsprechend John Searles Forderung nach einem philosophischen Rahmen für die Kognitionsforschung) sein sollen, so müsste sich zeigen lassen, dass (2) hilfreich für das Studium emotionaler Reaktionen ist. Die Wahrscheinlichkeit ist groß, dass dies nicht der Fall ist: *dass* es Interaktionen gibt, ist sowieso klar, und wie man sich der Frage nach den Interaktionen experimentell nähert, geht aus (2) nicht hervor. Geht man dagegen davon aus, dass die betrachteten Wechselwirkungen durch ein dynamisches System[38] beschrieben werden können, so wird man nicht nur die Wirkung Frustration → Aggression, sondern auch Rückkopplungen Aggression → Frustrationen in Rechnung stellen müssen. Wechselwirkungen ergeben sich als Resultat von simultan ablaufenden Aktivierungs- und Inhibierungsprozessen[39], weshalb das Merkmal der Kausalität der Beziehung zwischen funktionalen Größen an Relevanz verliert: zu jedem Zeitpunkt t gibt es einen Effekt Aggression → Frustration und Frustration → Aggression, und die Intensitäten dieser Wirkungen sind Funktionen der Zeit. Die Zustandsmenge ist nun auf einem Kontinuum definiert. Wollte man die Diskretheit dieser Menge beibehalten – vielleicht aus experimentalpsychologischen, beobacchtungstechnischen Gründen – so sollte man die diskrete Version als Approximation an die "wahre" kontinuierliche Version herleiten und dabei die Struktur der Wechselwirkungen beibehalten, um den Kern des Modells testen zu können. Eine Alternative bestünde darin, sich der beliebter werdenden Behauptung, mentale Ereignisse seien gar nicht auf neuronale Aktivitäten zurückführbar, zuzuwenden. Wem allerdings die Reduktion auf das Materielle unangenehm ist, kann sich denjenigen Anhängern der Philosophie des komputationalen Funktionalismus anschließen, die sich das Prinzip der multiplen Realisierbarkeit zu eigen gemacht haben, das das Tor zur Freiheit des Dualismus öffnet.

Man muß bedenken, dass die Turingmaschine gar nicht zur Beschreibung von Prozessen der hier betrachteten Art konzipiert worden sind, es geht bei Turingmaschinen darum, diskrete Folgen von Operationen abzubilden. Sie könnten also als *Modell* für gewisse Denkprozesse dienen, die sich als Folgen von Operationen darstellen lassen. Ob aber das Gehirn beim Denken tatsächlich im strengen Sinn diskrete Folgen erzeugt ist keineswegs klar. Die diskrete Repräsentation eines Denkprozesses als Folge von Operationen einer

[38] Gemeint sind Systeme, die durch Systeme von Differentialgleichungen beschrieben werden.

[39] Das allgemeine Prinzip wird z.B: in Murray (1989), insbesondere Abschnitt 5.4: Autocatalyses, Activation and Inhibition dargestellt.

Turingmaschine kann andererseits eine vernünftige Approximation des Denkprozesses darstellen, sofern eine gewissermaßen makroskopische Betrachtung eines Denkprozesses gewünscht ist[40]. Ein wesentliches Problem einer funktionalistischen Repräsentation der Dynamik mentaler Zustände liegt in der *a-priori*-Konzeption der Dynamik als einer diskreten Folge von Zuständen, und nicht als eine Approximation an den realen Prozess im neuronalen System. Sprevak (2017) kommt zu ähnlichen Schlüssen bezüglich der Rolle der Turingmaschine als Modell für unser kognitives System, und verweist auf Putnam (1975), der seinen eigenen Ansatz ebenfalls verworfen hat, wenn auch nicht wegen neuroanatomischer und neurophysiologischer Befunde über massive Parallelverarbeitung im Gehirn, wie sie u.a. von Edelman und Tononi (2004) vorgestellt werden. Es war die Einsicht in die phänomenologisch nicht weiter aufteilbare Struktur mentaler Ereignisse, die Putnam ur Abkehr von seiner Theorie bewogen. Sprevak argumentiert gleichwohl, dass die Turingmaschine vielleicht doch als Modell für eine "high-level"-Beschreibung kognitiver Aktivität dienen könnte. So kann man zur Modellierung von Problemlöseprozessen von der Annahme ausgehen, dass das Gedächtnis in ein Kurz- und ein Langzeitgedächtnis aufgeteilt wird, – ob es sich dabei um anatomisch verschiedene Lokalisationen handelt oder nicht, kann bei bestimmten Untersuchungen vernachlässigt werden, weil es bei einer gegebenen Fragestelllung nur darauf ankommmt, dass Informationen aus dem Kurzzeitspeicher mit einer gewissen Wahrscheinlichkeit ins Langzeitgedächtnis transformiert werden. Diese Hypothese ist funktionalistischer Natur, kann deshalb durch eine geeignet konzipierte Turingmaschine repräsentiert werden eben weil sie nicht explizit auf anatomische und neurobiologische Fragen eingeht. Der Wissenschaftlerin ist aber klar, dass es sich bei einer derartigen Beschreibung um eine herbe Vereinfachung handelt, die nur Sinn macht, weil man nur auf spezielle Aspekte der Gedächtnisspeicherung fokussieren will. Der Punkt dabei ist, dass es nicht die charakteristischen Eigenschaften einer Turingmaschine sind, die die Repräsentation eines Prozesses durch eine Turingmaschine für den Forschungsprozess nützlich machen, sondern es sind die psychologischen Annahmen, die im Zentrum der Forschung stehen; die Repräsentation durch Turingmaschinen ist dabei sekundär. Aus dem Konzept der Turingmaschine folgt noch nichts über

[40]Eine berühmt gewordene Aufgabe in der Denkpsychologie besteht darin, die Korrektheit einer Behauptung der Form $p \to q$, wobei p und q Aussagen sind, zu überprüfen. Versuchspersonen wrden vier Karten vorgelegt. Jede karte zeigt entweder einen Buchstaben oder eine Zahl, etwa "A", "D", "4", "7", und $p \to q$ ist die Aussage "Wenn eine Karte einen Vokal zeigt, so steht auf der Rückseite der Karte eine gerade Zahl". Die Aufgabe der Versuchsperson ist, diejenigen Karten umzudrehen, die einen Test der Behauptung ermöglichen. Dieses Experiment wurde zuerst von Wason (1968) beschrieben und wird bis heute diskutiert: die meisten Personen nennen die Karten "A" und "4", im Unterschied zur korrekten Lösung "A" und "7" ($p \to q$ impliziert nicht-$q \to$ nicht-p). Eine Diskussion des Denkverhaltens der Testpersonen im Rahmen der Theorie der Turingmaschinen erweist sich als wenig fruchtbar.

die Biologie des Gedächtnisses[41], s.a. auch Graves et al. 2014[42].

Eine kontinuierliche Dynamik läßt sich, eingebettet in milde Zusatzannahmen, im Rahmen der Theorie der Differentialgleichungen modellieren, und Systeme solcher Gleichungen lassen sich numerisch lösen, wenn keine analytische Lösung möglich ist, – was im Allgemeinen der Fall ist. Die Gleichungen reflektieren die Struktur der modellierten Dynamik, aber die Programme, mit denen man die Gleichungen numerisch löst, müssen keineswegs der modellierten Struktur in einer Eins-zu-eins-Weise entsprechen. Gedächtnisprozesse sind ihrer biologischen Struktur nach vermutlich kontinuierlich, aber man kann sie für bestimmte kognitive Prozesse, etwa Problemlöseprozesse, durch ein diskretisiertes System approximieren. Ein Mensch, der ein Problem löst, muß z.B. Zwischenergebnisse abspeichern. Hier kann man die Annahme machen, dass derlei Ergebnisse zunächst in einem Kurzzeitspeicher "abgelegt" werden, von wo aus sie mit einer gewissen Wahrscheinlichkeit in das Langzeitgedächtnis transferiert werden, so kann man den Gesamtprozess durch ein diskretisiertes Modell approximieren, dessen Dynamik durch ein Computerprogramm abgebildet werden kann, dass die Speicher- und Abrufprozesse simuliert, wobei das Programm keine Eins-zu-Eins-Kopie des simulierten Prozesses sein muß. Man sollte dabei die Rede von einem vom Langzeitspeicher getrennten Kurzzeitspeicher nicht allzu wörtlich nehmen, denn es ist ja denkbar, dass die Kurzzeitspeicherung nur eine vorlaufige oder unvollständige Art der Speicherung ist, die nicht an einen vom Langzeitspeicher getrennten Ort gebunden ist. Der gesamte Prozess läßt sich dann simulieren derart, dass die Ausgaben (outputs) des Programms auf den modellierten Prozess beziehen lassen. Die Beziehung zwischen dem modellierten Prozess und dem simulierenden Programm muß keineswegs eindeutig sein, es muß nur klar sein, wie die Ergebnisse des Programms auf den tatsächlichen Prozesse bezogen werden müssen. Das Programm könnte als funktionalistische Repräsentation der wahren Dynamik aufgefasst werden, die selbst keiner funktionalistischen Orgnisation im Sinne Putnams entspricht. Die Unterscheidung zwischen dem tatsächlichen Prozess und dessen funktionalistischer Simulation ist aber nicht Gegenstand der Theorie des Funktionalismus. Wenn man das eigentlich interessierende Modell mit den Resultaten der Computersimulation vergleicht ist die Struktur des simulierenden Programms von nachgeordneter Bedeutung. So könnte ein erfahrener Programmierer ein eleganteres, weniger Rechenzeit benötigendes Programm schreiben als ein Anfänger, etwa wenn "Funktionen" (Integrale, z.B. die Werte von Verteilungsfunktionen, Eigenvektoren und Eigenwerte von Matrizen, etc) numerisch berechnet werden müssen. Was wirklich interessiert ist die konti-

[41]"... there is no evidence in favour of the psychological or anatomical plausibility of an architectural distinction between storage an processing of information. Cordeschi & Frixione (2007), p. 38

[42]Pinna (2011), und Wells (1998) zur Fehlinterpretation von Turingmaschinen, wobei es aber umm eher extravagante Ansätze wie die *Extended Mind*-Theorie von Clark & Chalmers (1998) geht, auf die im Kapitel über den freien Will kurz eingegangen wird.

nuierliche Dynamik der Interaktionen, weniger die Programme, mit denen die Dynamik abgeschätzt werden soll.

Es sei noch darauf hingewiesen, dass mentale Zustände im Rahmen des Funktionalismus als Resultat einer Interaktion mit anderen mentalen Zuständen entstehen, ohne dass dabei auf eine immaterielle Kognitionssubstanz verwiesen wird. Gleichzeitig sollen aber wegen das MR-Prinzips bewußte Zustände und Zustandsinteraktionen nicht auf eine neuronale Basis reduzierbar sein, – die Rede ist dann von einem "nichtreduktiven Materialismus", bei dem keineswegs klar ist, was darunter verstanden werden soll, der oxymoronische Charakter dieses Ausdrucks läßt den Eindruck aufkommen, dass es sich um eine sprachliche Nebelkerze handelt. Eine andere Frage ist, woher die Maschinentabellen der postulierten Turingmaschinen kommen, die die kognitive Aktivität erklären sollen. Vielleicht existiert eine Super-Turingmaschine, die die Maschinentabelle für die Lösung eines speziellen Problems erzeugt. Wem das zu spekulativ ist kann empfohlen werden, auf Chalmers (1996) Conceivability-Argument zurückzugreifen: die Tatsache, dass man sich eine solche Super-Turingmaschine vorstellen kann (ohne dabei zu sehr ins Detail zu gehen) impliziert schon die Möglichkeit, dass sie existiert[43], weshalb sie dann auch existiert, – *in potentia*, aber immerhin. Ein nachgerade klassisches Argument gegen die Turingmaschine als allgemeines Modell für kognitive Aktivitäten ist, dass sie grundsätzlich seriell arbeitet, dass aber neuroanatomische und neurophysiologische Untersuchungen eher große Evidenz für massive Parallelverarbeitung bei der Informationsverarbeitung in realen Gehirnen geliefert haben (Edelman & Tononi (2009), Singer (2007)). Funktionalisten können allerdings darauf hinweisen, dass es mittlerweile auch parallelverabeitende Turingmaschinen gibt (Worsch (2020)). Man darf andererseits nicht vergessen, dass parallel verarbeitende Systeme auf einem normalen Computer, also einer Turingmaschine, simulieren kann, – ohne dass der Computer dabei zu einem parallelverarbeitenden System wird. Graves, Wayne und Danihelka (2014) haben ein neuronales Netzwerk konstruiert, das mit einem externen Gedächtnisspeicher verbunden ist, mit dem über Aufmerksamkeitsprozesse interagiert werden kann; sie konnten zeigen, dass das Modell im Prinzip einer Turingmaschine, bzw einer von Neumann-Architektur mit differenzierbaren Koeffizienten äquivalent ist, bei dem also kontinuierliche Zustandsübergänge möglich sind (Graves Wayne, Danihelka (2014)). Im hier diskutierten Zusammenhang ist es aber wichtig, zu sehen, dass bei Graves et al. nicht vom Konzept der Turingmaschine ausgegangen wurde, sondern die Äquivalenz nachträglich nachgewiesen wurde.

Kausalität, kausale Geschlossenheit: Die vorangehenden Betrachtungen sind angesichts der bereits vorhandenen Vorstellungen in der Psychologie eher banal, weshalb unklar ist, warum der Funktionalismus eine so bedeutende Rolle in der Philosophie des Geistes und in Teilen der Psychologie spielen konnte.

[43] Man kann sich ja auch vorstellen, dass der liebe Gott morgen das Gravitationsgesetz aufhebt, – also ist es möglich, dass es aufgehoben wird

Implizit wurde er schon immer angenommen, allerdings ohne dass Forderungen wie (2) explizit aufgestellt wurden. Dazu kommt die explizite Forderung, dass es sich bei der mentalen Dynamik um *kausale* Interaktionen handele. Diese Forderung ist, sofern man umgangssprachlich kommuniziert, trivial – man wird z.B. nach Dollard et al. (1939) aggressiv, *weil* man frustriert wurde –, andererseits ist sie fragwürdig, weil der Kausalitätsbegriff gewissermaßen naiv eingeführt wird: deutet man den Funktionalismus im Rahmen der Theorie dynamischer Systeme, so ist nicht mehr notwendig klar, was jeweils Ursache oder Wirkung ist: weil die verschiedenen Variablen, deren Dynamik betrachtet werden soll, direkt oder indirekt miteinander gekoppelt sind wirkt jede Variable auf jede ein und der Kausalitätsbegriff verliert seine erklärende Kraft. Der Zustand eines dynamischen Systems in einem Zeitpunkt t ist durch die Werte aller dynamiscch interagierenden Variablen des Systems zum Zeitpunkt t gegeben, er wird durch einen Punkt im n-dimensionalen Raum repräsentiert, und von Kausalität kann nur insofern geredet werden, als der Effekt aller dieser Variablen die Bewegung dieses Punktes durch diesen Raum steuert. Die Frage ist, ob die Beschreibung als *kausale* Wechselwirkung essentiell für die funktionalistische Konzeption ist. Putnam's "Aufweichung" der Kausalität durch Übergang zu einer probabilistischen Konzeption klärt in Bezug auf die Frage nach der Verursachung wenig bis nichts, weil das eben genannte Kopplungsprinzip ja erhalten bleibt. Im Übrigen impliziert der Sachverhalt, dass ein Prozess probabilistisch strukturiert ist, für sich genommen noch lange nicht, dass er akausal ist. wie die Betrachtungen zum Laplaceschen Dämon zeigen, ist es ja möglich, dass die kausalen Effekte einfach nicht berechenbar sind.

2.4.2 Lucas' Bedenken und Freys Theorem

Anders ist es vielleicht mit der grundsätzlichen Frage, ob ein mit dem Begriff der Turingmaschine so eng verknüpfter Ansatz wie der Funktionalismus nicht von vorn herein viel zu eng gewählt wurde. Wie schon angemerkt wurde wird im Rahmen des Funktionalismus auf die Frage nach der Rolle des Bewußtseins nicht weiter eingegangen. Der britische Philosoph John Lucas (1961) hat die Gödelschen Sätze als Ausgangspunkt für Argumentationen für die grundsätzliche Nichterklärbarkeit des Bewußtseins gewählt, wenn die Turingmaschine als kanonisches Grundmodell für alle Kognitionen angenommen wird. Lucas macht Gebrauch von dem Sachverhalt, dass Gödel gezeigt hat, dass jedes formale System einen Satz enthält, der im Rahmen des formalen Systems nicht bewiesen werden kann (dies ist der Satz (*) auf Seite 12). Der Satz ist aber in Gödels Theorem enthalten, was nach Lucas so gedeutet werden kann, dass Menschen – und nicht Wesen, deren kognitive Kompetenz im Rahmen dr Theorie der Turingmaschinen beschreibbar ist (s. das Halteproblem) – den Satz als wahr erkennen können. Lucas folgert, dass das kognitive System des Menschen nicht als formales System beschrieben werden kann.

Lucas' Argument liest sich zunächst sehr überzeugend, ist aber keineswegs unkritisiert geblieben. Die Debatte ist länglich, hält immer noch an (Lucas will einfach nicht aufgeben), und kann hier nicht in voller Breite wiedergegeben werden. Der Kern der Kritiken bezieht sich auf Ungenauigkeiten in den Lucasschen Argumenten, sowie ein etwas laxer Umgang mit dem zweiten Theorem des Gödelschen Theoreme, (2) (Benacerraf(1967)), demzufolge die für das Theorem 1 notwendige Bedingung der Konsistenz des formalen Systems nicht innerhalb eben dieses formalen Systems nicht bewiesen werden kann. Slezak (1984)hat die Kritik noch einmal vertieft, in dem er auf "subtil verteckte" Konfusionen in Lucas' Text verwies. Diese hängen u.a. mit dem Phänomen der Selbstreferenz zusammen. So argumentiert er (p. 23), dass die aus Gödels Theorem folgenden Beschränkungen einer Maschine sich auf sie, die Maschine, selbst beziehen, weil sie nicht imstande ist, eine selbstreferentielle Aussage zu "berechnen". Für Lucas sei es "äußerst trivial", dass er in Bezug auf die Maschine dieser Beschränkung nicht unterliegt. Die relevante Frage sei, ob Lucas imstande sei, seine eigene Gödel-Formel (eine wahre, aber in seinem Rahmen nicht beweisbare Aussage) aus *seinen* Axiomen und *seinen* Regeln zu beweisen. Das Problem sei nicht, was Lucas über eine Maschine wissen könne, sondern was er über sich selbst wissen kann.

Lucas hat ein intuitives Unwohlsein an rein funktionalistischen Theorien formuliert und mit immer neuen Argumenten auf die Gegenargumente reagiert, offenbar ohne die Kritiker überzeugen zu können. Dabei geht es den Kritikern weniger um eine Rechtfertigung des funktionalistischen Ansatzes, sondern um die Details des Gödelschen Theorems. Abgesehen davon bleiben die Fragen nach den Implikationen der Selbstreferenz erhalten, wie im Abschnitt über den freien Willen noch deutlich werden wird.

Es sei in diesem Zusammenhang aber auf ein Argument des österreichischen Philosophen Gerhard Frey[44] eingegangen, das in der Diskussion über die Möglichkeit von Theorien des Bewußtseins anscheinend übersehen wurde. Frey bezieht sich dabei nicht explizit auf den Ansatz, kognitive Prozesse anhand des Begriffs der Turing-Maschine zu modellieren, sondern auf die Tatsache, dass jede Theorie sprachlich formuliert werden muß. Sein Ansatz besteht in der Anwendung des cantorschen Diagonalverfahrens, das schon von Gödel und Turing zur Herleitung ihrer Theoreme (die Gödelschen Sätze bzw. das Turingsche Halteproblem) verwendet wurden.

Freys Argument Frey (1980) geht davon aus, dass die Existenz eines Gehirnmodells bedeutet, dass die "Funktionen und und Elemente [des Modells] in eine Beziehung zu den Funktionen und Elementen des Bewußtseins gesetzt werden können", wobei den Funktionen des Bewußtseins eindeutig denen des Modells entsprechen müssten, aber nicht notwendig auch umgekehrt: nicht alle Funktionen des Nervensystems haben Auswirkungen auf das Bewußtsein.

[44](1915 – 2002))

Weiter postuliert Frey, dass psychischen Akten und Relfexionen das "Wesen des Bewußtseins" ausmachen, die sich wiederum in unserer natürlichen Sprache abbilden, die selbst reflexiv ist. Wenn es ein Gehirnmodell gibt, so ist es auch beschreibbar, wobei die verwendete Sprache auch formale und bildhafte Elemente enthalten kann. Generell gilt ja, dass jeder Kalkül und jede formale Sprache nur formuliert werden kann, wenn er eine natürliche Sprache einbettbar ist. Jede Theorie des Bewußtseins enthält reflektive Aussagen, d. h. Aussagen über Aussagen. Ein Beispiel ist (i) die Aussage "ich denke", und die über diese Aussage (ii) "ich denke, dass mein Denken nicht konsistent ist". Die Aussage (ii) ist reflektiv. Eine Theorie des Bewußtseins sollte nur aus endlich vielen Aussagen bestehen, – würde diese Aussage nicht zutreffen, enthielte eine gegebene Theorie also aus mehr als endlich viele Aussagen, so wäre sie nutzlos: man müßte unendlich viele Aussagen lesen und verstehen, um die Theorie zu verstehen, was gegen den Sinn einer Theorie verstößt. Frey hat nun mittels des cantorschen Diagonalverfahrens, analog zum Vorgehen Gödels beim Beweis seiner Unvollständigkeitssätze, und Turings beim Beweis des Halte-Theorems nachgewiesen, dass keine endliche Theorie existieren kann (Abschnitt 3. Modell und Sprache, p. 69).

Die Gesamtheit der sprachlichen Ausdrücke werde mit S bezeichnet. Die den sprachlichen Ausdrücken entsprechenden Bewußtseinselemente und -funktionen werden mit B_s bezeichnet, und die den B_s entsprechenden Gehirnfunktionenen seien G_s. Frey macht die folgende

Annahme: S, B_s und G_s sind eindeutig aufeinander abbildbar. Die Gesamtheit aller Hirnfunktionen sei G; dann $G_s \subset G$.

Dann ergibt sich die Frage, ob G_s mit S beschrieben werden kann. Nimmt man einen Isomorphismus zwischen S und G_s an, so reduziert sich die Frage auf die, ob S mit S beschrieben werden kann und die Frage ist, ob sich die natürliche Sprache selbst beschreiben kann. Gödel (1931) hat jedenfalls gezeigt, dass sich ein Kalkül, also eine formale Sprache, nicht selbst beschreiben kann, während die natürliche Sprache ihre eigene Metasprache – also eine Sprache, die eine andere Sprache (die "Objektsprache") beschreibt – ist. Um zu zeigen, dass die sprachliche Beschreibung des Gehirnmodells möglich ist, muß Frey weitere Begriffe definieren:

Reduzible und nicht reduzible Reflexionsprädikate: Es gibt psychische Akte, die sich auf "äußere" Gegenstände richten; beziehen sie sich auf andere psychische Akte, so heißen sie *Reflektionen* (Frey, p. 19). In Sätzen werden Prädikate über das jeweilige Subjekt des Satzes ausgesagt, *reflexive Prädikate* haben als Subjekt einen ganzen Satz, sie drücken Urteile über Urteile aus bzw. ordnen ihnen Prädikate wie "wahr" oder "falsch" aus.

Zwei aufeinander folgende Reflexionsprädikate sind *reduzibel*, wenn sie zu einem einzigen solchen Prädikat zusammengefasst werden können: "Es ist wahr, dass A wahr ist" kann zu "A ist wahr" zusammengefasst werden. Nicht-

reduzibel sind "ich glaube", "du glaubst". a und b seien zwei Reflexionsprädikate. Dann kann man beliebige Reihen $abbaaba\ldots$ bilden In einem endlichen Text kommen stets nur endlich viel Reflexionsprädikate vor: Frey spricht von "finiten Reflexionsstrukturen". Dann kann man transfinite Reflexionen betrachten, die darin bestehen, über potentiell unendliche Reihen von reflexiven Prädikaten zu denken. Die Frage ist, ob derartige unendliche Reihen durch endliche formale sprachliche Mittel vollständig beschrieben werden können. Da die Sprache nur endlich viele Elemente (= Worte) enthält, kann es nur abzählbar viele solche Folgen geben, d.h. die Elemente der Folge können durchnummeriert werden, was bedeutet, dass jeder Folge genau eine natürliche Zahl zugeordnet werden kann; der Begrifff 'überabzählbar' wird im Anhang, Abschnitt 3 im Zusammenhang mit dem cantorschen Diagonalverfahren erläutert. Jedenfalls lassen sich die Folgen alphabetisch anordnen. Man hat dann

(1) $p_{11}, p_{12}, p_{13}, \ldots$
(2) $p_{21}, p_{22}, p_{23}, \ldots$
(3) $p_{31}, p_{32}, p_{33}, \ldots$
$\vdots$
(n) $p_{n1}, p_{n2}, p_{n3}, \ldots$

wobei p_{ij} eine Folge von a und b repräsentiert. Man nimmt dann aus der i-ten Reihe das i-te Element p_{ii}, für $i = 1, 2, \ldots$, und ersetzt dort a durch b und b durch a. Man erhält eine Reihe, die mit keiner der anderen Reihe übereinstimmt. Damit ist bewiesen, dass in jeder natürlichen Sprache mit wenigstens zwei nicht-reduzierbaren Reflexionsprädikaten in transfiniten Reflexionen unentscheidbare Sätze oder Paradoxien entstehen.

Daraus folgt, dass jede Beschreibung von S und G mittels S unvollständig ist, d.h. es gibt keine formale und vollständige Theorie des Bewußtseins, sofern diese als durch Aussagen über das Reflexionsvermögen definiert angenommen wird. Analoge Resultate auf der Basis von Turings Halte-Problem oder Gödels Theoremen werden oft als "limitierende Ergebnisse" bezeichnet: Die Gödel-Theoreme (Freys Argument kann man als eine Variante dieser Theoreme sehen, wobei Frey nicht auf die Bedingung der Konsistenz, also der Widerspruchsfreiheit, der jeweils betrachteten Theorie eingeht. Der Entwicklung der Mathematik und der Physik als einer durchmathematisierten Wissenschaft haben die Gödel-Theoreme nicht geschadet, und man kann davon ausgehen, dass neurowissenschaftliche Untersuchungen zur Genese des Bewußtseins ebenfalls kaum eine Einschränkung darstellen werden. Wie Frey selbst schreibt (p. 76) wird mit seinem Argument letzlich nur ausgesagt, dass das Bewußtsein nicht vollständig beschreibar ist, weil es selbst ein prinzipiell offenes System ist.

Eine allgemeine Diskussion der möglicherweise limitierenden Effekte der Gödelschen Theoreme hat Buechner (2010) gegeben. Die Details seiner Argumente sind zu technisch, um sie hier wiederzugeben – sie würden mindestens

ein ganzes Kapitel ausmachen, hier genügt eine Zusammenfassung. Die Gödel-Theoreme sind nur dann limitierend in dem Sinne, dass sie wissenschaftliche Ergebnisse ausschließen, wenn Ausssagen mit *mathematischer Gewißheit* gemacht werden sollen. Aber das ist normalerweise nicht der Fall, denn mathematische Modelle sind üblicherweise Approximationen und es immer wieder vorkommt, dass verschiedene Modelle für relevante Variablenbereiche zu sehr ähnlichen, empirisch nicht unterscheidbaren Voraussagen kommen. Dieser Sachverhalt gilt für alle Wissenschaften, und es ist nicht zu sehen, warum die Neuro- und Kognitionswissenschaften hier eine Ausnahme sein sollen.

2.4.3 Multiple Realisierbarkeit und Reduzierbarkeit

Turing (1937/1950) formulierte die These, dass (i) alle kognitiven Aktivitäten als komputationale Prozesse betrachtet werden können, und dass (ii) die komputationalen Prozesse durch Turingmaschinen abgebildet oder repräsentiert werden können. Diese These brachte Putnam und Fodor auf die Idee, dass alle kognitiven Aktivitäten multipel realisierbar und *deshalb* nicht auf biologische, letzlich also physikalische Prozesse reduzierbar seien. Natürlich sei es möglich, neurowissenschaftliche Untersuchungen über die Grundlagen kognitiver Aktivitäten durchzuführen, aber eigentlich seien derartige Untersuchungen überflüssig. Die Idee wurde eher beiläufig und ohne weitere Begründung vorgestellt, wurde und wird aber von Philosophen des Geistes bereitwillig aufgenommen (Koons & Bealer (2010)), wenn auch nicht von allen (Eliasmith (2002)).

Die Begriffe Multiple Realisierbarkeit (MR) und Reduzierbarkeit werden in der Philosophie des Geistes seit Putnams (1967) These, dass MR die Nichtreduzierbarkeit mentaler Ereignisse impliziert, zwar wie ein zueinander korrespondierendes Paar von Begriffen behandelt, sind aber eigentlich verschiedene Themenbereiche. Es zeigt sich, dass der Begriff der Reduktion einer wissenschaftlichen Theorie auf eine andere Theorie nicht so eng gefasst werden kann, wie es in den Argumentationen von Putnam und vor allem von Fodor (1974) unterstellt wird, und überdies folgt die Nichtreduzierbarkeit nicht aus der MR-These. Sie lässt sich statt dessen aus Davidsons These der Anomalität des Mentalen herleiten: Davidson argumentiert, dass mentale Ereignisse zwar mit gewissen physischen Ereignissen identisch seien, das Mentale aber insofern anomal sei, als die Beziehungen zwischen derartigen Ereignissn nicht durch strenge, physische Gesetze beschreibbar seien. Daraus folge, dass es keine psychophysischen Gesetze gebe: die Rede ist von der Nichtreduzierbarkeit des Mentalen auf das Physische und damit von Davidsons Begriff des nichtreduktiven Materialismus. Auch dieses Argument ist eher intuitiv, repräsentiert einen allenfalls diffus charaktierisierten Dualismus und ist auf jeden Fall nicht zwingend, worauf weiter unten noch explizit eingegangen werden wird. Zunächst wird auf das Konzept der multiplen Realisierbarkeit fokussiert.

2.4.3.1 Zum Begriff der Multiplen Realisierbarkeit In Abschnitt 2.3.1, Seite 22, wurde eine von Shagrir (1998) formalisierte Definition der Multiplen Realisierbarkeit (MR) vorgestellt, derzufolge die MR durch eine "wilde Disjunktion" – die Aussage (7), Seite 22 — charakterisiert ist. Es mag nützlich sein, Shagrirs Definition mit der ursprünglichen, von Putnam (1967) gegebenen Definition zu vergleichen, in der Form von Bickle (2020) (das Original wird hinzugefügt, um den Eindruck zu vermeiden, hier sei falsch übersetzt worden:[45]):

1. Zumindest einige mentale Zustände seien in verschiedenen physischen Substraten realisierbar, also *multipel realisierbar*,
2. Wenn ein mentaler Zustand S in verschiedenen physischen Substraten realisierbar ist, dann kann S nicht identisch mit einem speziellen physikalischen Substrat sein,
3. Folgerung: es existieren mentale Zustände, die nicht identisch mit irgendeinem physischen Substrat sind.

Bemerkenswert an dieser Spezifikation der MR ist, dass Punkt 1. nicht mehr hinreichend für die Definition des Begriffs der multiplen Realisierbarkeit ist, sondern dass die Punkte 2. und 3. hinzugefügt werden. Bickle merkt an, dass die Aussagen 1. bis 3. zumindest miteinander kompatibel sind, d.h. keine widerspricht einer den jeweils beiden anderen, so dass man sagen kann, dass das Konzept der MR zumindest intern konsistent ist. Das bedeutet nicht, dass die MR-These auch wahr ist. So kann bereits die Aussage 1. falsch sein. Aussage 2. verwundert, weil sie eine Aussage über eine Disjunktion, also eine oder-Verbindung ist, und Disjunktionen sind wahr genau dann, wenn mindestens eine der durch "oder" verknüpften Teilaussagen wahr ist, – also könnte auch die Aussage, dass mentale Ereignisse eine neuronale Basis haben, wahr sein, und 3. wird dann gegenstandslos. Akzeptiert man 1., so kann die Folgerung 3. als Stützung der These der Nichtreduzierbarkeit, also einer im Prinzip dualistischen Konzeption des Mentalen genommen werden. Damit ist kein Rückzug auf den Substanzdualismus nach Art René Descartes, der die Existenz einer speziellen Substanz, der *res cogitans*, als Trägerin mentaler Prozesse postulierte gemeint, wehalb gerne der oben genannte Begriff des *nichtreduktiven Physikalismus* angeführt wird, demzufolge das Mentale schon eine irgendwie physikalische Basis hat, aber gleichwohl nicht auf das "Physikalische" reduziert werden kann. Vielleicht ist ein mentaler Äther gemeint, von dem man nicht weiß, woraus der besteht, die Hauptsache ist, dass er er nicht physikalisch ist. Das Argument der Wilden Disjunktion ist ebenfalls icht überzeugend. God-

[45] Stated in canonical form, Putnam's original multiple realizability argument draws an anti-identity theory conclusion from two premises: (i) (the multiple realizability contention) (At least) some mental kinds are multiply realizable by distinct physical kinds, and (ii) if a given mental kind is multiply realizable by distinct physical kinds, then it cannot be identical to any one (of those) specific physical kind. Then one has the the anti-identity thesis conclusion: (At least) some mental kinds are not identical to any one specific physical kind.

bey (1978) argumentiert, dass es einfach als Implikation der Putnamschen Behauptung, dass mentale Ereignisse funktional definiert werden können, gesehen wird. Godbey gibt eine knappe Beschreibung dieser Deduktion (p. 433): Jede Anzahl von physikalischen Zuständen kann einen funktionalen Zustand "realisieren", – Marsbewohner können das, ebenso Computer. Wenn wir Schmerz empfinden, so kann das an der Aktivierung bon C-Fasern liegen. Wenn aber der Marsbewohner oder der Computer Schmerz empfinden, so kann es nicht an den C-Fasern liegen, weil sie diese nicht haben. Daraus kann man folgern, dass ein mentaler Zustand nur mit einer Disjunktion von physischen Zuständen identisch sein kann, wobei ein physikalischer Zustand existiert, der das mentale Ereignis tatsächlich erzeugt. Deswegen kann die Disjunktion nicht aus endlich vielen möglichen Zuständen bestehen, sondern muß unbegrenzt sein. Daraus folge aber auch das Argument der Nichtreduzierbarkeit: man könne sich zwar vorstellen, dass man alle möglichen physikalischen Zustände aufzählen könne, die mit einem bestimmten mentalen Ereignis korrelieren, aber diese Aufzählung würde kaum ein psychogisches Gesetz ergeben.

Godbey erinnert nun daran, dass eine 'erfolgreiche' Reduktion wie die der Thermodnamik auf die statistische Physik – also letztlich auf die newtonsche Dynamik – kaum möglich gewesen wären, wäre Ludwig Boltzmann in dieser Weise vorgegangen. Wärme kommt in verschiedenen Substanzen vor, die aus organischen oder anorganischen Stoffen bestehen, egal, ob sie bei uns auf der Erde oder auf dem Mars lagern. Was, so könnte man fragen, haben sie denn gemein, um das Phänomen 'Wärme' zu zeigen, dass sie also ein physikalisches Merkmal haben? Folgte man der Fodorschen Argumentation, so wäre die Thermodynamik nicht reduzierbar. Aber sie ist reduzierbar. Dies liege, so Godbey, an der Klassifizierbarkeit der Objekte: in der Physik könne man Objekte nach ihrer Masse, ihren Geschwindigkeiten und ihrer Beschleunigung klassifizieren, und relativ zu bestimmten Klassifikationen existieren Gesetze, – bei gleichzeitiger Ko-Extensivität zu irgendwelchen unbegrenzten Disjunktionen.

Das Wesentliche bei einer Wissenschaft sei aber, so Godbey, die richtige Wahl der Variablen. In der newtonschen Physik (hier die Mechanik), sei es die Masse, die Geschindigkeit und die Beschleunigung eines Körpers, nicht aber sein Geruch, seine Farbe, und was sonst noch für Eigenschaften eines Körpers notiert werden könne. Es sind in der Tat diese Größen, die die Reduktion der Thermodynamik auf die statistische Physik ermöglichen. Mit großer Wahrscheinlichkeit müssen die Psychologie die für die Reduktion relevanten Variablen noch gefunden werden; es werden nicht nur die Spike-Folgen von Neuronen sein. Das MR-Argument ("the same functional state (mental state) can be realized by practically anything", wie Fodor 1974, p. 18, schreibt) reicht nicht aus, um die Nichtreduzierbarkeit der Psychologie zu beweisen. Ähnliche Argumente findet man z.B. bei Francescotti (1997) und Clapp (2001)).

Die Frage nach der Reduzierbarkeit einer wissenschaftlichen Theorie auf eine basalere Theorie ist nicht nur in Bezug auf die Psychologie von Inter-

esse, so dass in Abschnitt 2.4.3.4 ein Blick auf den allgemeinen Begriff der Reduzierbarkeit geworfen werden soll.

Zunächst stellt sich die Frage, ob die Behauptung, mentale Ereignisse seien multipel realisierbar, überhaupt sinnvoll ist. Der Hinweis auf unterschiedliche Hirnstrukturen ist in dieser Abstraktheit, d.h. ohne Diskussion von Daten, bemerkenswert leer. Auch die zur Stützung der MR-These vorgebrachte Behauptung, die Lokalisation von aktivierten Hirnregionen für bestimmte mentale Ereignisse könne variieren, weshalb mentale Ereignisse nicht auf die Aktivität bestimmter neuronaler Populationen zurückgeführt werden können, impliziert ja noch lange nicht, dass ein mentales Ereignis keine neuronale Basis hat, – der Befund signalisiert ja nur, dass man eine Erklärung für die Invarianz des Mentalen gegenüber dieser Variation finden muß. Die These der multiplen Realisierbarkeit innerhalb ein und desselben Gehirns unterstellt, dass die Reduktion *nur* mentaler auf neuronale Ereignisse nur nach Maßgabe der Identitätstheorien von Smart und Place gedacht werden kann, – was aber mit den neueren Ergebnissen der Hirnforschung nicht kompatibel ist (Singer (2007)[46]). Was die MR zwischen Species angeht kann auf die Phylogenese verwiesen werden, – schließlich stammen wir von Meererstieren ab und grundsätzliche Mechanismen könnten sich früh herausgebildet haben und dann vererbt und verfeinert bzw. modifiziert worden sein. Es kann auch sein, dass es spezifische, an die jeweilige Umgebung angepasste Entwicklungen gegeben hat, was aber zu der Vermutung führen kann, dass die Repräsentation der Umgebung für einen Octopus anders als die für einen Hai oder für einen Menschen. Dann gilt die MR-These nicht in der in den Punkten 1. bis 3. behaupteten Allgemeinheit.

Bechtel & Mundale (1999) haben die MR-These in Bezug auf wesentliche Resultate der Hirnforschung diskutiert. Die metaphysische These sei, dass mentale Prozesse einfach die durchgeführten Operationen selbst seien, ohne dass irgendeine Basis dafür existieren müsste. Darüber hinaus unterliegt der Pro-MR Argumentation die Annahme, dass ein mentales Ereignis M ohne weitere Differenzierung über verschiedene Spezies, aber auch über verschiedene Zustände eines Individuums als eben das Ereignis M konzipierbar sei. Schmerz ist für Primaten - natürlich einschließlich des Menschen – wie für den Octopus eben Schmerz, ebenso der Hunger, etc. Da sich die Gehirne der verschiedenen Lebewesen aber voneinander unterscheiden wird philosophischerseits gefolgert, dass derartige mentale Ereignisse unabhängig vom Gehirn existieren. Dies soll auch für kleine Veränderungen ein und desselben Gehirns gelten. Bechtel et al.

[46]"The brain is a highly distributed system in which numerous operations are executed in parallel and that lacks a single coordinating center. This raises the questions of i) how the computations occurring simultaneously in spatially segregated processing areas are coordinated and bound together to give rise to coherent percepts and actions, ii) how signals are selected and routed from sensory to executive structures without being confounded, and finally iii) how information about the relatedness of contents is encoded. One of the coordinating mechanisms appears to be the synchronization of neuronal activity by phase locking of self-generated network oscillations." Singer (2007) p. 1

sprechen deshalb von der unterschiedlichen Körnigkeit der Begriffe, das Mentale werde gewissermaßen grobkörnig charakterisiert, während Hirnzustände "feinkörnig" diskutiert werden. Das ist ein interessanter Sachverhalt und die Reaktion eines empirisch orientierten Menschen ist vermutlich, ihn näher zu untersuchen. Zumindest einige Philosophen sehen hier aber schon eine Bestätigung der MR-These, was wiederum Bechtel et al. dazu bringt, von einer philosophischen Fiktion zu sprechen. Der Erfolg der MR-These in der *Philosophy of Mind* (PoM) sei auf ein kontextuales Vakuum zurückzuführen: es werde kein Zusammenhang angegeben, der auf die Frage, ob zwei gleiche mentale Ereignisse dieselben organischen Zustände (Hirnzustände) implizieren, eine Antwort habe.

Brodmann[47] lieferte eine erste Kartierung des Gehirns, wobei er komparativ arbeitete: 55 Spezies mit allein 11 verschiedene Arten von Säugetieren mit dem Ziel, *homologe Bereiche* bei den verschiedenen Tierarten zu identifizieren. Sein Hauptziel war es, Gemeinsamkeiten zwischen den Species zu finden. Andere Forscher, z.B. Heinrich Vogt[48] kartierten feiner, dh mit höherer Auflösung. Spätere Forscher untersuchten insbesondere das visuelle System, etwa Felleman & van Essen (1991); sie identifizierten 32 verschiedene, untereinander verlinkte Bereiche, in denen visueller Input verarbeitet wird. Sie differenzierten in Bezug auf drei Kriterien: Architektonik, Konnektivität, topographische Organisation. Brodmanns Fokus auf Architektonik erwies sich nur für eine Minderheit der Bereiche des visuellen Kortex als nützlich, etwa 50 % der Bereiche wurden durch topographische Organisation, d.h. die Projektion des visuellen Feldes für jeden Bereich determiniert, während die Konnektivität für nahezu alle Bereiche von Bedeutung ist (Mishkin et al. (1983)). Diese Kartierungen dienen der Charakterisierung funktionell relevanter Gebiete. Generell gilt, dass bereits beim primären Bewußtsein komplexe Wechselwirkungen zwischen neuronalen Teilpopulationen eine Rolle spielen, die mit den simplizistischen Vorstellungen eines Hirnzustandes, von denen viele Philosophen auszugehen scheinen, wenig zu tun haben. Philosophische Betrachtungen sind, in Bezug auf philosophische Thesen, eher verifikationistisch, indem sie Gründe für eine Bestätigung der MR-These suchen, – etwa kleine Abweichungen von Vorhersagen im fMRI (Beckermann (2008)). Einen guten Einblick für eine neurowissenschaftlich fundierte Diskussion liefern Edelmann und Tononi (2004); eine Darstellung ihrer Befunde sprengt den hier gegebenen Rahmen. Philosophische Betrachtungen allein werden das Rätsel der Bewußtwerdung nicht lösen können. Kim (2002) räumt ein, dass Philosophen, die von der Gültigkeit der MR-These ausgehen, vergessen haben, dass derartige Thesen auch getestet werden müssen[49]. Für empirisch arbeitende Kognitions- und Neuro-

[47] Korbinian Brodmann (1868 – 1918) Neuroanatom und Psychiater

[48] 1875 – 1957), Neurologe und Psychiater

[49] Empirisch arbeitenden Wissenschaftlern ist das sowieso klar, aber unter Philosophen ist anscheinend eine Ausnahme.

wissenschaftler haben die philosophischen Thesen dementsprechend allenfalls den Rang von Hypothesen. Kim kritisiert wiederum Bechtel & Mundale, weil sie zu sehr auf die Homologie von Hirnarealen fokussieren, – die Homoplasie sei wichtiger (Flügel von Vögeln und Fledermäusen haben keine gemeinsamen Vorfahren, sondern haben sich separat entwickelt). Eine analoge Aussage gilt für die Augen von Primaten und Octopussen (vergl. Couch (2004) für eine detaillierte Diskussion).

2.4.3.2 Eliasmith-Kritik Funktionalisten nehmen an, dass jeder mentale Zustand als Beziehung zwischen Inputs, Outputs und sie verbindende mentale Zustände begriffen werden kann (Putnam (1975)). Dies bedeutet, dass jede kognitive Funktion komplett durch "high-level" Beschreibungen, unabhängig von der Implementation charakterisiert werden kann. Zwei verschiedene Implementierungen können funktional äquivalente Beschreibungen haben. Damit wird gewissermaßen unter der Hand der Begriff der 'funktionalen Äquivalenz' eingeführt. Eliasmith (2002) will zeigen (Abschnitt 4), dass diese Konzeption von Äquivalenz nicht möglich ist. Er zitiert zunächst Ned Block (1980, p. 178), demzufolge diese Äquivalenz so offensichtlich wahr sei, dass sie als Grundposition bzw. Axiom für die Philosophie des Geistes gewählt werden könne (Eliasmith, p. 7):

> die Behauptung [der Äquivalenz] hat eine so große *prima facie*-Plausibilität, dass die Last des Beweises beim Kritiker liegt, der Gründe beibringen muß, um seine Zeifel zu begründen[50].

Zusammen mit der These der Multiplen Realisierbarkeit (MR-These) ergibt sich dann die (Hypo-)These, dass alle mentalen Prozesse durch Turingmaschinen repräsentiert werden können, – ganz unabhängig von der biologischen, allgemein der physischen Realisierung. Eliasmith (2002) fasst die Argumentation in sieben Punkten zusammen:

1. Systeme mit Mentalität ("mind") sind kognitive Systeme.
2. Kognitive Systeme sind komputationale Systeme.
3. Turingmaschinen beschreiben jedes komputationale System vollständig.
4. Deshalb können Turingmaschinen jedes kognitive System vollständig beschreiben.
5. Turingmaschinen sind unabhängig von ihrer Implementation definiert (d.h. sie sind *funktional* definiert.
6. Deshalb kann jedes kognitive System unabhängig von seiner Implementation definiert bzw. charakterisiert werden.
7. Also können Systeme mit Mentalität unabhängig von ihrer Implementatierung definiert bzw charakerisiert werden.

[50] ... the claim has such overwhelming prima facie plausibilty that the burden of proof is on the critic to come up with reasons for thinking otherwise.

Eliasmith will zeigen, dass die Implementierung eines kognitiven Systems und und seine Funktion nicht so einfach voneinander getrennt werden können wie dies mit der MR-These behauptet wird; die MR-These sei weder testbar noch aufschlußreich, – sie sei schlicht inhaltlos.

Zunächst die These 2, dass alle kognitiven Systeme komputationale Systeme sind. Axiom 3 gilt nur, wenn angenommen wird, dass der Begriff des komputationalen Systems identisch mit dem des Systems von Komputationen ist, die als Menge von Input-Output-Transformationen konzipiert ist. Turingmaschinen beschreiben Komputationen, aber nicht physische Systeme, die Berechnungen durchführen. In Aussage 2 wird der Term 'komputationales System' gebraucht, um sich auf Apparate oder Vorrichtungen zu beziehen, da reale kognitive Systeme physische Systeme sind, während Systeme von Komputationen dies nicht sind. Man hat es also mit einer Mehrdeutigkeit des Begriffs 'komputationales System' zu tun, *die das Argument* als ungültig ausweist.

1963 hat der russische Mathematiker Andrei Nikolajewitsch Kolmogorov (1903 – 1987) den Begriff der Algorithmischen Komplexität, auch einfach Kolmogorov-Komplexität[51] eingeführt. Es sei zunächst daran erinnert, dass ein Programm stets als Folge von Nullen und Einsen dargestellt werden kann. Eine solche Folge heißt auch *Binärsequenz*. Die Kolmogorov- bzw. Algorithmische Komplexität ist dann durch die Aussage

> Die *Algorithmische Komplexität* einer unendlichen Binärsequenz s ist die länge des kürzesten Programms, das s ausgibt.

definiert. Denker & leCun (1993) haben diesen Begriff auf einen Nenner gebracht, aus dem Eliasmith (2002) eine grundsätzliche Kritik am Konzept der Multiplen Realisierbarkeit und damit am Funktionalismus herleitete. Es wird zunächst das hier relevante Resultat von Denker & leCun vorgestellt. Ein Programm wird in einer festgelegten Programmiersprache geschrieben. Sofern es irgendwann stoppt, gibt es eine Binärsequenz s aus. Das Programm entspricht einer Turingmaschine, die auf einem leeren Band gestartet werden. Die *Länge* eines Programms ist die Anzahl von Bits[52]. Zur Illustration betrachte man einige Folgen: (a) eine Folge von 10 Millionen Nullen (sehr einfach), (b) eine Folge von Folgen der Art 01, die 500 000 Male wiederholt werden (fast so einfach wie (a), (c)). Eine Folge von Nullen und Einsen, die durch das Werfen einer Münze (Kopf = 1, Zahl, = 0) entsteht (ziemlich komplex), (d) die erste Million Dezimalzahlen der Zahl π. Diese Folge ist kaum komplexer als die Folgen (a) und (b), weil jede Dezimalzahl deterministisch vorausgesagt, d.h berechnet werden kann.

Die algorithmische Komplexität K einer Folge von Symbolen S ist nun

[51] Solomonoff, Chaitin

[52] Ein Bit ist die Information, die z.B. bei einem Münzwurf gewonnen wird. (Hoffmann (2011), p. 316)

(s. oben) gleich der Länge des kürzesten Programms, das für einen gegebenen Computer $c1$ diese Folge generiert und dann stoppt. Formal läßt sich wie in

$$K_{c1} := \min_{P}[\text{Len}(P)|\text{eval}(c1, P) = S] \tag{8}$$

ausdrücken, wobei $\text{Len}(P)$ die Länge des Programms ist und $\text{eval}(c1, P)$ ist das Resultat des Programms P auf dem Computer $c1$. K_{c2} ist analog dazu die algorithmische Komplexität für einen zweiten Computer $c2$, der der Folge S eine andere Komplexität zuordnet. Kolmogorov zeigte, dass für einen universellen Computer $c1$ und einen beliebigen Computer $c2$ die Beziehung

$$K_{c1} \leq K_{c2}(S) + k_{12} \tag{9}$$

gilt, wobei k_{12} eine von $c1$ und $c2$ abhängende, aber nicht von S Konstante ist. k_{12} ist die Länge eines *Emulatorprogramms* $EM(c1|c2)$, das für $c2$ benötigt wird, um $c1$ zu *emulieren*[53], dass als $c1$-Programm auf $c2$ nachbildet. Dividiert man die Gleichung (9) durch K_{c2}, so erhält man

$$\frac{K_{c1}}{K_{c2}} \leq 1 + \frac{k_{12}}{K_{c2}} \tag{10}$$

Offenbar gilt

$$\lim_{K_{c1}(S)\to\infty} \frac{k_{12}}{K_{c2}} = 0,$$

da k_{12} eine Konstante ist, so dass man

$$\lim_{K_{c1}(S)\to\infty} \frac{K_{c1}}{K_{c2}} = 1. \tag{11}$$

Dies bedeutet, dass die Konstante k_{12} für wachsenden Wert von $K_{c1}(S)$ vernachlässigt werden kann und die Abhängigkeit der Komplexität von der Wahl der Computer $c1$ und $c2$ vernachlässigbar wird, so dass K ein allgemeingültiges Maß für Komplexität wird; wird die Differenz $K_{c1} - K_{c2}$ betrachtet kann allerdings die Konstante k_{12} nicht vernachlässigt werden.

Die Komplexität ist also nur im Grenzfall (11) für verschiedene Computer vernachlässigbar, aber dieser Grenzfall liegt im Allgemeinen, d.h. für endliche Folgen, nicht vor, zumal Nebenbedingungen für das Verhalten zu berücksichtigen sind: Organismen müssen auf Herausforderungen mit hinreichender Geschwindigkeit reagieren können. Dieser Sachverhalt hat Konsequenzen für die Annahme der multiplen Realisierbarkeit. Der Punkt ist ja, dass es hier nicht darauf ankommt, dass verschiedene Computer alle dieselbe Klasse von Funktionen berechnen können, sondern dass ihr Aufbau ihre Komplexität im Kolmogorovschen Sinne beeinflusst. Je höher die Komplexität, desto mehr Zeit

[53]Emulator: ein System bzw. Programm, dass ein anderes System oder Programm aspektweise simulieren kann.

wird für die Berechnung einer gegebenen Funktion benötigt. Im allgemeinen "äquivalente" (d.h. sie berechnen dieselbe Klasse von Funktionen) Computer sind deshalb nicht "gleich": "Werden identische Turingmaschinen implementiert, so sind die entsprechenden Komplexitäten im Allgemeinen nicht identisch" (Eliasmith, p. 9). Eliasmith illustriert diesen Sachverhalt an einem einfachen Beispiel: ein Organismus muß entscheiden, ob ein anderer Organismus "Freund" oder "Feind" ist. Eine Turingmaschine, die diesen Entscheidungsprozess charakterisiert, bestimmt für sich genommen noch nicht das Verhalten, es kommt zusätzlich auf die Art der Implementierung der TM an: eine Implementierung, die 10 Sekunden für die Entscheidung benötigt, dürfte das Überleben mit größerer Wahrscheinlíchkeit sichern als eine Implementierung, die 10 Minuten für die Entscheidungsbildung benötigt. Jedes psychische System kann in Termen von nahezu unendlich vielen verschiedenen virtuellen Maschinen beschrieben werden, ohne das entschieden werden könnte, welche davon explanatorisch relevant sind. Die Turingmaschine sagt also nur wenig über das kognitive Verhalten eines Organismus aus, die Komplexität der Implementierung muß ebenfalls berücksichtigt werden. Damit sind aber die für den Funktionalismus wesentlichen Grundannahmen (3) bis (6) falsch, und damit ist auch die Forderung, dass "high-level"- Wissenschaften wie die Psychologie nicht reduzierbar und damit "'methodologisch autonom" seien hinfällig.

Was also ist das beste Argument für den Komputationalismus? Nach van Gelder (1995) ist es das "What else?"-Argument: das einflußreichste Argument der Anhänger des Komputationalismus sei die Behauptung, es gäbe einfach keine Alternative[54]. Dass Argumente dieser Art allgemein für Neuro- und KognitiionswissenschaftlerInnen überzeugend sind ist wenig wahrscheinlich.

2.4.3.3 Neuroplastizität und Multiple Realisierbarkeit Es ist schon angemerkt worden, dass der Begriff der Multiplen Realisierbarkeit (MR) das Tor zum Dualismus ist, und wer einmal zum Dualismus gefunden hat, scheint nicht mehr davon lassen zu wollen; Putnam (1975) folgert dementsprechend, dass MR die Nichtreduzierbarkeit der Psychologie auf physische Prozesse impliziere. NeurowissenschaftlerInnen neigen allerdings zu empirischen Tests, von denen einige hier berichtet werden sollen. Wie Maimon & Hemmo (2022) ausführen bedarf es dazu scharfer Definitionen, insbesondere des Begriffs der MR[55]. Eine Definition ist bereits durch die "wilde Disjunktion" (7), Seite

[54] As Allen Newell (1990) recently put it: "...although a small chance exists that we will see a new paradigm emerge for mind, it seems unlikely to me. Basically, there do not seem to be any viable alternatives. This position is not surprising. ... In: "Are There Alternatives?" in W. Sieg, ed., Acting and Reflecting (Boston: Kluwer, 1990).

[55] "for something to be multiply realized, it must be simultaneously the same, yet different, on two different levels. Concerning the mental and the physical, for multiple realizability to uphold, sameness is required at the level of the mental, concurrent with a difference at the physical level."... "the difference must be such that it is *relevant to their performing the same function the differences among would-be realizers must be 'other' than mere individual*

22) gegeben worden, für die Zwecke der folgenden Argumentionen soll sie aber noch einmal spezifiziert werden. So sei M ein mentales Ereignis und φ_j, $j = 1, \ldots, n, \ldots$ seien physische Prozesse, $\varphi_j \neq \varphi_k$ für $j \neq k$, die mit mentalen Ereignissen $M_j(\varphi_j)$ einhergehen, wobei insbesondere

$$M = M_j = M_k \quad \varphi_j \neq \varphi_k \; j \neq k \tag{12}$$

gelte, d.h. die φ_j erzeugen ein und dasselbe M. Dann ist M mehrfach (multipel) realisierbar. Diese Definition entspricht den Definitionen von Putnam (1975) und Fodor (1974, 1997). Die Idee ist, dass verschiedene physische ("neurobiochemische") Prozesse zu ein und demselben mentalen Ereignis M führen, – bloße Ähnlichkeit der M_j genügt nicht, um von MR zu sprechen, denn der Fall $M_j \neq M_k$ für $\varphi_j \neq \varphi_k$ ist ja mit einer Identitätstheorie kompatibel. Ließe man, was eigentlich vernünftig wäre, zufällige, kleine Unterschiede zwischen M_j und M_k zu, um bei der MR-These bleiben zu können, müsste man Kriterien für die Bedeutung von "klein" einführen, oder sogar Signifikanztests, – das Klammern an der MR-These würde immer akrobatischer werden. Denn der Fall $M_j \neq M_k$ für $\varphi_j \neq \varphi_k$ wäre ja mit einer Identitätstheorie kompatibel, auch wenn der Unterschied zwischen M_j und M_k klein und dementsprechend die beiden Ereignisse einander ähnlich sind. MR muß deswegen so radikal definiert werden, denn dass ähnliche – aber nicht identische – physische Prozesse ähnliche mentale Ereignisse erzeugen können würde in der Neurowissenschaft kaum bezweifelt werden, und man müsste sich bei neurowissenschaftlichen Untersuchungen mit der Frage abgeben, wie unähnlich mentale Ereignisse sein müssen, damit nicht mehr von MR gesprochen werden kann.

Wie Maimon & Hemmo ausführen, liefert das Phänomen der Neuroplastizität für die Anhänger der MR-These die beste Evidenz für MR (s. a. Endicott (1993), p. 312). Sie argumentieren dann, dass das Gegenteil der Fall sei: entgegen der in der Philosophie des Geistes allgemein akzeptierten Sichtweise verweisen empirische Daten auf "type-type", also $(M - \varphi)$-Korrelationen, die moderne Varianten der Identitätstheorien stützen. Es beginnt damit, dass verschiedene Typen von Neuroplastizität unterschieden werden müssen: (i) die strukturelle, und (ii) die funktionale Neuroplastizität. Bei der strukturellen Neuroplastizität handelt es sich primär um synaptische Plastizität. Damit ist die Verstärkung odere Schwächung synaptischer Verbindungen zwischen Neuronen gemeint. Diese Veränderungen von synaptischen Verbindungen bilden die Basis von Lern- und allgemein von Gedächtnisprozessen und wurde schon von dem Neurophysiologen Donald Hebb (1949) diskutiert; ihm wird die Redeweise "neurons that fire together wire together" zugeschrieben, die in der Neuro- und Kognitionswissenschaft offiziell als *Hebbs Regel* bekannt ist[56]. Der

difference the variation must not merely map onto individual differences (Polger and Shapiro 2016, 67), zitiert nach Maimon & Hemmo (2022), p. 109)

[56]"Let us assume that the persistence or repetition of a reverberatory activity (or "trace") tends to induce lasting cellular changes that add to its stability. ... When an axon of cell A

MR-These entsprechend müssen irgendzwei verschiedene physische Prozesse ("physical types" in der Sprache der Philosophy of Mind) ein und dasselbe mentale Ereignis ("mental kind") generieren, um von MR reden zu können. Das ist bei der strukturellen Neuroplastizität aber gerade nicht der Fall, weil neuro-biochemische Prozesse ja gerade die Veränderung der mentalen Ereignisse bedingen; Maimo & Hemmo (2022), p. 113, liefern biochemische Details dieser Prozesse.

Die funktionale Neuroplastizität wird durch Veränderungen der physiologischen Aspekte der Funktionsweise der Zelle definiert, also durch Veränderungen der Rate, mit der Aktionspotentiale ("spikes") erzeugt werden, oder der Wahrscheinlichkeit, mit der chemische Signale ausgelöst werden. Diese bewirken Veränderungen der synaptischen Verbindungen, oder erhöhen die Synchronizität bei der Bildung von Spike-Folgen. Aus Sicht der Philosophie stützen diese Prozesse die MR-These, weil angenommen wird, dass jeweils dasselbe mentale Ereignis erzeugt wird. Eine genauere Betrachtung zeigt aber, dass diese Annahme nicht in der behaupteten Allgemeinheit gemacht werden kann.

So betrachten Maimo er al. Prozesse oder Zustände, die durch Schlaganfälle ausgelöst werden, etwa Arten der Aphasie. Aphasie ist, einer Charakterisierung Damasios (1992) entsprechend, "die Unfähigkeit, mentale Repräsentationen in Sprache zu übersetzen, und umgekehrt." Sie resultiert aus Schäden der Sprachzentren auf der linken Hemisphäre. Es zeigt sich, dass derlei Schädigungen zu kompensatorischen Prozessen an homologen Positionen in der rechten Hemisphäre führen. Die rechte Hirnhälfte übernimmt Funktionen, die vor der Schädigung ausschließlich in der linken Hemisphäre ablaufen konnten. Diese Übernahme kann, als eine erste Reaktion, als Nachweis einer multiplen Realisierbarkeit gedeutet werden, weil nun derselbe mentale Zustand durch zwei verschiedene physische Prozesse erzeugt werde. Allerdings ist es so, dass ein Schlaganfall eine Folge von Prozessen auslöst: zum einen die Aktivierung von neuronalen Netzwerken auf der linken Seite, zum anderen eine *Disinhibierung* von homologen Bereichen in der rechten Hemisphäre, und darüber hinaus miteinander konkurrierende Prozesse bezüglich der hemisphärischen Dominanz. In der akuten Phase unmittelbar nach dem Schlaganfall gibt es in beiden Hemisphären keine Reaktionen auf sprachbezogene Aufgaben. In der folgenden subakuten, bis zu sechs Monate dauernden Phase existieren neuroplastische Aktivitäten, und in der rechten Hemisphäre können sprachbezogene Aktivitäten beobachtet werden. Wie Saur et al (2006), p. 1371, feststellen erfolgt die Regenerierung der sprachlichen Kompetenzen "in bereits existierenden, bilateralen Netzwerken von Neuronen mit einer Hochregulierung der nicht geschädigten Bereiche und der Rekrutierung von perilateralem Gewebe und homologen rechten Sprachbereichen" (Maimon et al., p. 114). Maimon et

is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased." Hebb (1949),p. 62

al. erinnern nun an Putnams und Fodors Argument, dass von MR dann gesprochen werden könne, wenn die Adaption in verschiedenen homologen Bereichen durch *verschiedene* physische Prozesse, die keine gemeinsamen Komponenten haben, erfolgen muß. Es ist aber so, dass die hier diskutierten mentalen Prozesse, d.h. die sprachlichen Abläufe, keineswegs identisch sind, d.h. die Prozesse in der rechten Hemisphäre sind im Vergleich zu den ursprünglichen, in der linken Hirnhälfte lokalisierten Abläufen ausgesprochen begrenzt, wohingegen die korrespondierenden neuro-biochemischen Prozesse *nicht* unterschiedlich sind. Die Prozesse in der rechten Hirnhälte sind Prozesse der Disinhibierung von bereits existierenden neuronalen Verbindungen, und *diese* Mechanismen existieren sowohl zur Zeit vor der Schädigung sowie in der subaktiven Phase. Daraus folgt, dass bei der Adaptation homologer Bereiche nicht von multipler Realisierbarkeit geredet werden kann.

Kompensatorische Maskerade ist ein neuroplastischer Prozess, bei dem geschädigtes neuronales Gewebe, das die Ausübung bestimmter kognitiver Aktivitäten erschwert oder unmöglich macht, gewissermaßen umgangen wird. Dabei wird gesundes neuronales Gewebe rekrutiert, um diese kognitive Aktivitäten zu ermöglichen. Die Rede ist von *periläsionaler Aktivität*[57]. Ist die Schädigung die Folge eines Schlaganfalls, so beginnt die *chronische Phase* ca sechs Monate nach dem Anfall: die Regeneration stabilisiert sich auf einem bestimmten Niveau, wobei es zu einer *Deaktivation* der korrespondierenden Position in der rechten Hemisphäre kommt. Wenn sich die linke Hemisphäre nicht hinreichend erholt bleibt es in der chronischen Phase bei einer aktivierten rechten Hälfte und einer Behinderung der Sprachfunktion(Fischer-Baum et al. (2017). Man könnte diese periläsionaler Aktivität, also Relokation der relevanten Aktivität, als Manifestation der multiplen Realisierbarkeit sehen, d.h. als Bestätigung der MR-These. Dazu muß aber gezeigt werden, dass die "neuen" neuronalen Prozesse die relevanten neuro-biochemischen Prozesse *nicht* mit den ursprünglichen Mechanismen teilen. Es zeigt sich aber, dass die Sprachfunktion im Laufe der chronischen Phase wieder von der linken Hemisphäre übernommen wird und dabei *derselbe* neuro-biochemische Zustand wieder hergestellt wird. Wenn es zu einer unvollständigen Regeneration kommt sind die kognitiven Prozesse nicht identisch mit denen, die vor der Schädigung abliefen. Zusammengefasst bedeutet dies, dass sich die MR-These nicht aufrechterhalten lässt (Maimon & Hemmo (2022), p. 116).

Intermodale Plastizität. Gegeben seien zwei Neuronenpopulationen N_1 und N_2, wobei N_1 für die Verarbeitung von Signalen aus einer Sinnesmodalität M_1 und N_2 für die Verarbeitung von Signalen aus einer anderen Sinnesmodalität M_2 zuständig ist. Intermodale Plastizität bedeutet, dass z.B. N_1 die Signale für M_2 verarbeitet. Die Rede ist auch von *sensorischer Substitution*; sie wird ebenfalls als Beweis für die MR-These betrachtet.

[57] periläsional - um die (= peri) Läsion, also den Schaden, herum

Es zeigt sich aber, dass diese Interpretation der sensorischen Substitution zu pauschal ist und die tatsächlichen Prozesse nur ungenau abbildet. So kann man blinde und sehende Personen miteinander vergleichen. Bei einer sehenden Person wird der visuelle Kortex ausgehend von den Retinae über den optischen Nerv stimuliert. Für den Blinden gibt es diese Stimulierung nicht, sie erhalten Informationen über die Umwelt über andere Sinnesquellen, u.a. durch taktile Stimulation. Ein Beispiel hierfür ist das "Lesen" der Braille-Schrift. Dabei müssen taktile Muster diskriminiert werden. Wenn eine blinde Person Braille-Schrift liest findet man auch eine Aktivierung von Neuronen im visuellen Kortex, während sich beim taktilen Lesen von Braille-Schrift bei einer sehenden Person eine *Deaktivierung* visueller Neuronen zeigt (Sadato, 1996, 2005a, 2005b). Man kann nun sehende Versuchspersonen (Vpn) mit verbundenen Augen Braille-Schrift lesen, d.h. mit den Fingerspitzen fühlen lassen. Nach ein paar Tagen zeigten sich in ihrem visuellen Kortex die gleichen Aktivierungen, die bei blinden Vpn gefunden werden. Diese Aktivierungen verschwanden allerdings sofort, nachdem man den sehenden Vpn die Binde von den Augen genommen hatte. Die Geschwindigkeit dieses Stops der Aktivierung ist nicht mit der Möglichkeit einer Neubildung von Verbindungen kompatibel und zeige, so Maimon & Hemmo, dass der Wechsel in den Aktivierungen eher das Resultat der Aufhebung einer Inhibierung bereits existierender Verbindungen, bzw. einer Stärkung von *bereits existierenden Verbindungen* sei. Dies bedeute, dass der Okzipitallappen und der primäre visuelle Kortex sowohl bei sehenden wie auch bei blinden Personen multimodal sei und deswegen sowohl visuelle Bilder wie auch räumliche Vorstellungen generieren könne (Pascual-Leone & Hamilton (2001))

Damit ergibt sich für Vertreter der MR-These ein Dilemma (Maimon & Hemmo, p. 118):

(a) die empirischen Daten legen nahe, dass die Aktivierung des visuellen Kortex das Korrelat der visuellen Erfahrung (des mentalen Ereignisses) ist. Diese Annahme ist gleichbedeutend mit der Akzeptanz einer Identitätstheorie. Ist man nun ein Anhänger der MR-These, so könne man diese Konsequenz nicht akzeptieren und sei gezwungen, die Annahme, dass die visuelle Erfahrung kortexbasiert ist, aufzugeben.
(b) Eine alternative Möglichkeit ist, zu akzeptieren, dass blinde und sehende Braille-Leser dieselben visuellen Erfahrungen haben. Dieser Schuß widerspricht der MR-These, die man nun fallen lassen muß, – statt dessen akzeptiert man die Aussage, dass die visuelle Erfahrung von blinden und sehenden Braille-Lesern letztlich auf denselben neuro-biochemischen Mechanismen der Aktivierung des visuellen Kortex[58] beruht. Dieser Sachverhalt stützt eine allgemeine Identitätstheorie und nicht die MR-These.

[58]"... mediated through the neural networks for tactile-visual cross-modal integration", Sadato (2005), p. 581

Maimon & Hemmo kommen zu dem Schluß, dass die empirischen Daten insgesamt gegen die MR-These sprechen. Bereits Bickle (2003) hatte argumentiert, dass es bestimmte molekulare Prozesse sind, die psychologischen Ereignissen unterliegen[59], d.h. die MR-These werde eben *nicht* durch synaptische neuroplastische Prozesse gestützt, sondern das zu einem mentalen Ereignis korrespondierende physische Ereignis sei ein molekularer Mechanismus, der "unterhalb" dieser Prozesse ablaufe. Hemmo, & Shenker (2015) argumentieren dementsprechend für eine Neufassung der Identitätstheorie ("brain-state hypothesis", p. 120), die sie als "flat physicalism" bezeichnen. Dabei handelt sich um eine Typ-Typ-Identitätstheorie, die analog zu einer verallgemeinerten Version der Thermodynamik entwickelt wurde. Eine Darstellung dieser Theorie sprengt aber den hier gegebenen Rahmen, so dass auf Details nicht eingegangen werden kann. Während Neuroplastizität gewisse neurologische Strukturen voraussetzt, besagt die *brain state*-Theorie, dass eine Eins-zu-eins-Beziehung zwischen biochemischen und mentalen Prozessen existiert[60].

Rewiring cortex 1999 publizierte eine Gruppe von Neurowissenschaftlern ein aufsehenerregendes Experiment: es war ihnen gelungen, bei Frettchen visuelle Information, die normalerweise von der Retina über den Thalamus in den visuellen Kortex geleitet wird, vom Thalamus aus in den auditiven Kortex umzuleiten (Sur et al. (1999)) und andere). Dann wurden visuelle Stimuli gesetzt, so dass die resultierende retinale Aktivität in den auditiven Kortex geleitet wurde. Der zentrale Befund dieser Untersuchungen ist, dass die Frettchen lernten, mit dem auditiven Kortex zu sehen. Diese Befunde wurden zunächst so interpretiert, dass die Aktivierung des auditiven Kortex das Resultat einer "Neu-" oder "Umverdrahtung" ("rewiring") sei, – ein Ausdruck, der wohl aus der Welt der Elektronik entliehen wurde. Vertreter der MR-These sehen diese Daten als Bestätigung der MR-These, da anscheinend ein mentales Ereignis – eine visuelle Wahrnehmung – durch eine andere Neuronenpopulation als die des visuellen Kortex generiert wurde.

Sharma et al. (2000) leiteten bei neu geborenen Frettchen Teile der neuronalen Verbindung von der Retina zum visuellen Kortex (V1) vom corpus geniculatum ab und in den Bereich A1 des auditiven Kortex um. Nach Erreichen des Erwachsenenstatus wurden die Frettchen unter Verwendung bildgebender Verfahren ("optical imaging") getestet: visuelle Stimuli[61] erzeugten im primären auditiven Kortex A1 orientierungsabhängige Aktivitäten analog zu den orientierungspezifischen Zellen im visuellen Kortex V1. Die horizontale Vernetzung im A1 entsprach ebenfalls der in V1, obwohl diese Vernetzung

[59] "shared molecular mechanisms realize shared psychological features and processes', Bickle (2003), p. 157

[60] ... they support the "brain-state hypothesis," namely, that the same kinds of biochemical processes in the brain are correlated with the same kinds of mental states", p. 123

[61] "rectangular gratings", d.h. Muster aus abwechselnd hellen und dunklen Streifen mit unterschiedlichen Orientierungen

nicht das Resultat spezieller Manipulierung war. Dieser Befund zeigt, dass die kortikalen Netzwerke durch sensorische Aktivierung strukturiert werden.

In einer analogen Untersuchung von von Melchner et al.(2000) wurden die Befunde von Sharma et al. nicht nur bestätigt, sondern noch differenziert. Die visuelle Information wurde nicht nur in den A1 umgelenkt, sondern man hatte zusätzlich im LGN[62] Läsionen gesetzt, so dass die Verbindung zum visuellen Kortex unterbrochen wurde. Die Details des Versuchsplans und der Daten müssen hier übergangen werden, eine wesentliche Schlußfolgerung ist jedenfalls, dass die sensorischen Kortexe über eine gemeinsame, basale Fähigkeit verfügen, auf Signale aus den jeweils anderen Bereichen zu reagieren. von Melchner et al. schreiben

> "... Die intrinsische Verarbeitung im primären auditorischen Kortex könnte in gewisser Hinsicht der Verarbeitung im primären visuellen Kortex ähneln. Diese Ähnlichkeit könnte den auditorischen Kortex in die Lage versetzen, visuelle Informationen zu verarbeiten. Eine plausible Erklärung für unsere Ergebnisse ist, dass primäre Bereiche des sensorischen Neokortex unabhängig von der Modalität bestimmte ähnliche, stereotype Operationen mit dem Input durchführen."[63] (p. 875)

Das heißt, dass z.B. der visuelle und der auditive Kortex einige basale Strukturen miteinander teilen. Es sind *diese* Strukturen und nicht die Umverdrahtungen, die die Reaktionen aus dem jeweils anderen sensorischen Bereich erzeugen (Makin et al. (2023), p. 6). Dieses Ergebnis spricht eindeutig gegen die MR-These: es ist *ein* Mechanismus, der die experimentellen Daten erzeugt[64], (vergl. Rabinowitz et al., 2011; Tian et al., 2013).

2.4.3.4 Zum Begriff der Reduktion Reduzierbarkeit heißt im Prinzip die Vereinheitlichung einer Wissenschaft. Man kann eine Theorie der Wärme formulieren, bei der 'Wärme' eine phänomenologische Größe ist, für deren Ausbreitung beschreibende Gleichungen aufgestellt werden können, ohne dass gesagt wird, welche physikalischen Prozesse "Wärme" ausmachen. Die Frage nach diesen Prozessen stellt sich gleichwohl, und dem österreichischen Physiker Ludwig Boltzmann (1844 – 1906) gelang es, das Phänomen Wärme auf die Bewegung von Teilchen, also etwa von Gasmolekülen und damit auf die Statistische Mechanik zurückzuführen: Temperatur entspricht der mittleren

[62] LGN = Lateral Geniculate Nucleus

[63] ... intrinsic processing in primary auditory cortex may be similar in certain respects to that in primary visual cortex. This similarity might allow auditory cortex to process visual information; indeed, a parsimonious explanation of our results is that primary areas of sensory neocortex perform certain similar, stereotypical operations on input regardless of modality."

[64] "the two brain areas share similar basic architecture, that is, feed forward thalamocortical inputs to layer IV that is then amplified by recurrent excitatory networks in cortex, and modulated by lateral inhibition, which underlies topographic representations"

kinetischen Energie der Teilchen. Diese Reduktion (Rückführung), bedeutet eine Vereinheitlichung der Physik: die Wärmetheorie kann damit letzlich auf die newtonsche Mechanik zurückgeführt werden. Das Ziel ist dabei nicht, die Thermodynamik überflüssig zu machen, sondern eine Erklärung für phänomenologische Gesetze zu finden. Um Grundlagenprobleme der Mathematik zu lösen haben etwa Gottlieb Frege (1848 – 1925) sowie Bertrand Russell (1870 – 1972) und Alfred North Whitehead (1861 – 1947) versucht, die Mathematik auf die Logik zurückzuführen, womit die Mathematik ein Teil der Logik geworden wäre[65]. Natürlich war es nicht das Ziel dieses Versuchs, die Mathematik überflüssig zu machen, vielmehr ging es um die Klärung der Frage, was Mathematik letzlich ausmacht. Auch die Chemie, die Biologie und viele andere Naturwissenschaften, die im Kern auf der Physik beruhen, werden wegen dieser Reduktion nicht überflüssig.

Im Falle der Reduktion der Psychologie auf die Neurobiologie (und damit auf die Physik) erhält man eine Erklärung mentaler Ereignisse in Termen neurobiologischer, biochemischer bzw. biophysikalischer Prozesse, und das Ziel ist ebenfalls *nicht*, die "high-level"-Wissenschaft Psychologie überflüssig zu machen, sondern zu einem tieferen Verständnis beobachteter Gesetzmäßigkeiten zu erhalten, – auch wenn es sich bei diesen "nur" um Ceteris-paribus-Gesetze (c.p.-Gesetze) handelt. In der Physik wird unter Reduktion oft nur die Deduzierbarkeit einer Theorie als Spezialfall einer allgemeineren Theorie verstanden (Berry (1994)). So läßt sich die Newtonsche Mechanik als Spezialfall der Speziellen Relativitätstheorie (SRT) darstellen: wenn die Geschwindigkeit v eines Körpers hinreichend klein im Verhältnis zur Lichtgeschwindigkeit c ist, d.h. wenn $\delta = v/c \to 0$, d.h. wenn δ "sehr klein" wird. In diesem Fall geht die SRT in die newtonsche Mechanik über. Im Beispiel der Thermodynamik als Spezialfall der Statistischen Mechanik hat man $\delta = 1/N \to 0$, was als Idealisierung für den Fall einer sehr großen Anzahl N der Partikel aufzufassen ist[66]. Offenbar besteht der Reduktionsprozess in einer Herleitung des Spezialfalls aus einer allgemeineren Theorie, weshalb die Begriffe *Korrespondenz* und *Kommensurabilität* notwendige Bedingungen für die Reduktion spezifizieren sollen: zwei Theorien korrespondieren zueinander, wenn die eine Theorie als Spezialfall der allgemeineren deduziert werden kann, und die Theorien sind kommensurabel, wenn sie logisch kompatibel sind.

Fodor (1974) basierte seine Behauptung, dass mentale Phänomene nicht auf physikalische Prozesse reduziert werden können, auf einer von Nagel (1961) formulierten Theorie der Reduktion; auf die Nagelsche Theorie wird in Abschnitt 2.4.3.5, Seite 55, noch ausführlich eingegangen. Die Struktur der Re-

[65] Dieses Projekt scheiterte, u.a. an Cantors Paradox (Cantor, G. (1899) "Beiträge zur Begründung der transfiniten Mengenlehre", und das eng damit verwandte Russellsche Paradox (1901)– Die Menge aller Mengen, die sich nicht selbst als Element enthalten

[66] Diese Darstellung ist skizzenhaft, eine ausführliche Darstellung sprengt den Rahmen dieses Textes.

duktion einer Theorie auf eine andere war natürlich schon vor Nagel (1961) Gegenstand wissenschaftstheoretischer Diskussionen, und Schaffner (1967) hat bereits vier verschiedene Auffassungen des Reduktionsbegriffs diskutiert, die kurz vorgestellt werden sollen, um die Besonderheit von Nagels Theorie evaluieren zu können:

Das NWQ-Paradigma: Hier steht **N** für Nagel (1960, 1961) W für Woodger (1952) Q für Quine (1964). Schaffner nennt dieses Paradigma *direkt*: die Basisterme der einen Theorie werden auf die Basisterme einer zweiten Theorie bezogen, wobei die Axiome und Theoreme der reduzierten Teorie aus denen der reduzierenden Theorie abgeleitet werden. Dabei kann es passieren, dass Ausdrücke der reduzierten Theorie nicht in der reduzierenden Theorie vorkommen: so gibt es den Ausdruck "Gen" nicht in der organischen Chemie. Diese müssen dann in der reduzierenden Theorie besonders eingeführt werden. Das klassische Beispiel für diese Art der Reduktion ist Boltzmanns Reduktion der Thermodynamik auf die statistische Mechanik.

Das KO-Paradigma: Dieses Paradigma geht auf Kemeny & Oppenheim (1965) zurück; Schaffner spricht von *indirekter* Reduktion, weil man nicht einfach eine Theorie T_2 auf eine Theorie T_1 reduziere, sondern identische, beobachtbare Vorhersagen aus beiden Theorien herleitet. Schaffner nennt als Beispiel Lavoisiers Oxydationstheorie für den Verbrennungsprozess, die alle *beobachtbaren Fakten* vorhersagt, die auch von der Phlogistontheorie erklärt werden, d.h die Phlogistontheorie wird *erklärt*, wobei aber der Begriff 'Phlogiston' nicht in Termen der Oxydationstheorie erklärt wird.

Das PFK-Paradigma: Hier steht P für Popper[67], F für Feyerabend (1962), und K für Kuhn (1962). Diesen Autoren ging es um die Reduktion früherer auf spätere Theorien, wobei die ältere Theorie ein Spezialfall der neueren Theorie ist. Hier wird eine Theorie T_2 nicht aus der Theorie T_1 in einem formalen Sinn abgeleitet, sondern erklärt, warum die ältere Theorie als Erklärung funktioniert hat und nun durch die neuere korrigiert wird. Die ältere Theorie T_2 wird also nicht aus der neueren Theorie T_1 im strengen Sinn deduziert. Als Beispiel nennt Schaffner die Galileische Theorie des freien Falls, die auf die Newtonschen Axiome der Mechanik *und* auf das Newtonschen Gravitationsgesetz zurückgeführt wird (s. Kapitel II, Seite 43). Das Galileische Gesetz erweist sich als nicht aus der Newtonschen Physik herleitbar, statt dessen kann eine etwas kompliziertere Theorie deduziert werden, die Daten vorhersagt, die den aus Galileis Gesetz vorhergesagten entsprechen: *dieselben* Vorhersagen ergeben sich nur unter der Annahme, dass der Erdradius unendlich ist. Die reduzierte Theorie ist also nur eine Approximation. Der Sachverhalt der Approximati-

[67] Popper, K.R. (1957) The aim of Science, *Ratio*, I, 24–35

on erzeugt Komplikationen für eine allgemeine Analyse der Theorienreduktion.

Das Suppes-Paradigma: Dieses Pradigma geht auf Suppes (1957) zurück, der Reduktion nach Maßgabe eines Repräsentationstheorems für Modelle einer Theorie charakerisiert, worauf gleich noch eingegangen wird. Suppes bezieht sich dabei explizit auf die Reduktion der Psychologie auf die Physiologie: die sollte möglich sein, wenn es für jedes Modell einer psychologischen Theorie ein isomorphes Modell innerhalb der Physiologie existiert.

Diese Charakterisierung des Reduktionsbegriffs setzt den Begriff des Modells einer Theorie voraus. Modelle einer Theorie sind vereinfachende Spezifikationen einer Theorie, die oft für den Zweck des Tests spezieller Hypothesen konzipiert werden. Ein (stark vereinfachendes) Beispiel ist die Theorie, dass das Lernen durch einen probabilistischen Transfer vom Kurzzeit-ins Langzeitgedächtnis beschrieben werden kann: ein Wortpaares (Tisch – table) wird zunächs im Kurzzeitspeicher "abgelegt"' und von dort mit einer bestimmten Wahrscheinlichkeit ins Langzeitgedächtnis transferiert. Überdies wird z.B. angenommen, dass dieser Prozess unabhängig von den vorangegangenen Darbietungen des Paares abläuft. Der Sinn eines Modells besteht darin, spezielle Annahmen über den Prozess testbar zu machen. Experimente müssen dann so geplant werden, dass die Korrektheit der Annahmen getestet werden kann, – wobei sich zeigen kann, dass einige der Annahmen mit den zur Verfügung stehenden Mitteln gar nicht testbar sind. Der Vorteil eines modellorientierten Vorgehens besteht darin, den empirischen Gehalt einer Theorie abschätzen zu können.

Ney (2006) gibt ebenfalls eine knappe Einteilung von Theorien zur Reduktion: mit T werde die zu reduzierende Theorie bezeichnet (T von *target*, Ziel), mit B die Theorie (B Basis), auf die T reduziert werden soll. Dann können drei Typen von Reduktionstheorien betrachtet werden:

(a) *Das Übersetzungsmodell:* Alle wahren Aussagen – einschließlich der Gesetze – von T können in die Sprache der Theorie B übersetzt werden,
(b) *Das Ableitungsmodell:* Alle Gesetze von T können aus den Gesetzen von B abgeleitetet werden,
(c) *Das Erklärungsmodell:* Alle Beobachtungen, die durch T erklärt werden, können auch durch B erklärt werden.

Diese Modelle schließen einander nicht notwendig aus, sondern können sich wechselseitig ergänzen. Das allgemeine Ziel der Reduktion einer Theorie auf eine andere ist wieder die Einheit der Wissenschaft.

2.4.3.5 Fodors Reduktionsmodell Der vorangegangene Abschnitt illustriert die Komplexität des Reduktionsbegriffs, – um also eine Nichtreduzierbareit des Mentalen auf das Physische zu begründen, muß man erklären, warum

man welchen Reduktionsbegriff zugrundelegt, denn offenbar folgt nicht aus allen Konzeptionen der Theorienreduktion die behauptete Nichtreduzierbarkeit. Fodor legt seinen Betrachtungen den von Nagel (1961) spezifizierten Begriff zugrunde, ohne diese Wahl weiter zu begründen. Nagels Betrachtungen bezogen sich auf Reduktionen, wie sie in der Physik vorgenommen worden waren und deren Ziel insbesondere die Erklärung von Sachverhalten war, und so leitete er seinen Reduktionsbegriff aus der *deduktiv-nomologischen Theorie der Erklärung* ab, der zufolge die Gesetze einer Theorie aus denen einer basaleren Theorie logisch deduziert werden sollen. Carnap et al (1975)[68] hatte gefordert, dass dabei zueinander korrespondierende Begriffe aus beiden Theorien identifiziert werden müssen, was durch die Formulierung von *Brückengesetzen* geleistet werden könne. Dementsprechend illustrierte Nagel seine Ideen u.a. am Beispiel der Thermodynamik, die auf die kinetische Gastheorie reduziert werden kann. Die Abbildung des Temperaturbegriffs der Thermodynamik auf den der mittleren kinetischen Energie der Gasmoleküle ist ein Beispiel für ein Brückengesetz. Fodor hat die Struktur einer Reduktion (so, wie er sie versteht) schematisch dargestellt: S_i sei ein beschreibendes Prädikat einer speziellen Wissenschaft (etwa der Psychologie), und P_i sei ein "physikalisches" Prädikat, also ein Prädikat aus einem Teilgebiet der Physik (oder, allgemeiner, aus einer "lower level" Wissenschaft, – relativ zur betrachteten "high-level" Wissenschaft). Mit "$\rightarrow$" werde eine Kausalitätsbeziehung bezeichnet, "$\leftrightharpoons$" steht für 'überbrücken'. Für alle Gesetze von S der Form

$$S_1 x \leftrightharpoons S_2 x \tag{13}$$

(Wenn x das Prädikat S_1 hat, genau dann hat x auch das Prädikat S_2) soll nun gelten, dass Brückengesetze der Form

$$S_1 x \leftrightharpoons P_1 x \tag{14}$$

$$S_2 x \leftrightharpoons P_2 x \tag{15}$$

gelten, ebenso ein Gesetz

$$P_1 x \rightarrow P_2 x \tag{16}$$

(alle x mit dem Prädikat p_1 haben auch das Prädikat p_2) derart, dass (13) eine deduktiv gültige Folgerung von (14), (15) und (16) ist. In der Nagelschen Theorie der Reduktion repäsentiert (13) ein Gesetz aus der Thermodynamik, (14) und (15) sind Brückengesetze und (16) ist ein Gesetz der statistischen Mechanik.

Soll nun die Reduktion den Physikalismus rechtfertigen, so bedeutet (Fodor (1974), Seite 100 oben) $\leftrightharpoons$ nichts weniger als "contingent event identity"[69]: (14) bedeutet dann "Jedes Ereignis, das aus x-en besteht, die S_1 erfüllen,

[68] Carnap, R., Neurath, O.: Wissenschaftliche Weltauffassung - der Wiener Kreis, in: Schleichert,H.(Hg.), Logischer Empirismus - der Wiener Kreis, München: 1975,201-222

[69] It is widely held that if an object a is identical (or non-identical) to an object b, then

ist identisch mit einem Ereignis, das aus x-en besteht, die P_1 erfüllen, und umgekehrt". Ein Beispiel für ein Brückengesetz – wieder aus der Physik – ist die Aussage, dass wenn (i) ein Gas die Temperatur T hat, so hat (ii) das Gas die mittlere molekulare kinetische Energie $\bar{E}_{kin}$. Wenn diese Aussage mit der These gekoppelt wird, dass jedes Ereignis, das unter ein Gesetz einer speziellen Wissenschaft fällt ebenfalls einem Brückengesetz wie (14) oder (15) genügt, dann drückt es die Allgemeinheit der Physik in Bezugauff eine spezielle Wissenschaf´t aus (Fodor (1074), p. 100 oben).

Diese Interpretation wäre dann ein Beispiel für eine "klassische Reduktion". Die Relation $\leftrightharpoons$ wird von Fodor als eine Bikonditionalität gedeutet. Angewandt auf die Frage nach einer Reduktion des Mentalen auf das Physische liefert das Prinzip der Bikonditionalität aber einen Widerspruch zur These der Multiplen Realisierung, die ja gerade besagt, dass wegen

$$S_1 x \leftrightharpoons P_1 x \vee P_2 x \vee \cdots \vee P_n x \tag{17}$$

ein mentales Ereignis auf ganz verschiedene Weise generiert werden kann. Deshalb könne, so die Fodorsche Lehre, das Mentale eben nicht auf bestimmte neuronale Ereignisse zurückgeführt werden. Zum Beispiel: x glaubt, dass Schnee weiß ist und die Annahme der Reduzierbarkeit bedeutet, dass x in einem Hirnzustand b_1 ist, wenn x ein Mensch ist. ist x aber ein Schimpanse, so ist x in einem Zustand b_2, und wenn x Roboter ist, so ist x in einem elektronischem Zustand e, und wenn schließlich x ein Alien aus einer fernen Galaxie ist, so ist x in einem Zustand g, d.h. ein bestimmter Zustand grünen Schleims. Nach Ansicht Fodors und seiner Anhänger ist die Disjunktion, also der Ausdruck auf der rechten Seite von (17), kein Prädikat irgendeiner Wissenschaft, so dass der gesamte Ausdruck (17) kein Gesetz sei.

Man muß an dieser Stelle darauf hinweisen, dass Reduktion in der Physik darin besteht, eine Theorie von größerer Allgemeinheit zum Zweck der Erklärung auf eine eingeschränktere Theorie zu reduzieren: die statistische Physik Boltzmanns auf die Thermodynamik, die Relativitätstheorie Einsteins auf die newtonsche Physik, etc. Dabei streben gewisse Parameter der allgemeineren Theorie gegen einen Grenzwert, – in der statistischen Physik strebt $\delta = 1/N$ wegen $N \to \infty$ gegen Null, in der Speziellen Relativitätstheorie strebt der Faktor $\gamma = 1/\sqrt{1 - v^2/c^2}$ wegen $v \to 0$ gegen 1 (s. oben). Die Reduktion besteht gewissermaßen aus der Betrachtung asymptotischer Werte von Parametern. Eine derartige Betrachtung liegt bei der Fodorschen Diskussion der Reduzierbarkeit von "higher-level" Wissenschaften wie der Psychologie

it is necessary that a is identical (non-identical) to b. This view is supported an argument from Leibniz's Law and a popular conception of *de re modality*. On the other hand, there are good reasons to allow for contingent identity. Various alternative accounts of de re modality have been developed to achieve this kind of generality, and to explain what is wrong with the argument from Leibniz's Law. Schwarz, W. (2013). Contingent identity. *Philosophy Compass*, 8(5), 486-495.

auf "lower-level"-Wissenschaften wie die Neurobiologie nicht vor, es gibt keinen Parameter δ, der gegen einen Grenzwert strebt, wenn psychologische auf neurobiologische Gesetzmäßigkeiten zurückgeführt werden sollen.

Methodologische Autonomie der Psychologie (Fodor): Fodor geht insbesondere davon aus, dass für die Kognitionswissenschaft das Prinzip der Multiplen Realisierbarkeit gilt. Es wird postuliert, dass es demnach keine mentalen Ereignisse gibt, die bestimmten physischen Ereignissen zugeordnet werden können, so dass das für die Reduktion benötigte Prinzip der Bikonditionalität nicht gelten könne. Daraus wiederum folge, so Fodor, dass die Psychologie eine eigenständige Wissenschaft sei, d.h. die Psychologie sei *methodologisch autonom*. Das Standardbeispiel ist die Schmerzempfindung, die nicht nur bei Menschen, sondern auch bei Tieren, vom Affen bis zum Octopus oder noch niederen Tieren auftreten könne. Menschen und die einzelnen Tierarten unterscheiden sich aber hinsichtlich ihrer Hirnstrukturen, woraus folge, dass die Schmerzempfindung nicht auf die Aktivität eines Gehirns zurückführbar sei. Skeptische Hinweise auf die phylogenetische Entwicklung des Gehirns und die Tatsache, dass Empfindungen wie Schmerz durch Aktivierung von Neuronenpopulationen in sehr basalen Hirnregionen erzeugt werden scheinen keine argumentative Wirkung auf Autoren zu haben, die diese Argumention vertreten. Auf die Frage, ob das Argument der multiplen Realisierung eher sanguinische Leichtigkeit der Gedankenführung oder beklemmende Enge eines vom Wunsch nach Dualismus gelenkten Assoziierens (man möchte einen "Beweis" für eine Art von Dualismus finden) reflektiert wird weiter unten eingegangen.

Die frühen Formulierungen von Reduktionstheorien, wie sie im Logischen Empirismus enwickelt wurden und auf die sich Nagel (1961) beziehen sind natürlich kritisch diskutiert worden, schon weil der Begriff der Theorie, wie er von den Logischen Empiristen konzipiert wurde, nicht akzeptiert werden konnte. Fodor geht gar nicht auf diese Arbeiten ein. In Nagels (1961) Modell spielt der Begriff des Brückengesetzes eine wesentliche Rolle. Dieser Begriff wird wiederum in Termen der in (17) angegegebenen, durch das Symbol $\leftrightharpoons$ angezeigten Bikonditionalität formuliert, den Fodor für seine Art des Nachweises für die Nichtreduzierbarkeit der Psychologie auf die Neurobiologie benötigt.

2.4.3.6 Hookers Theorie der Reduktion Bickle (1992a) argumentiert, dass grundsätzlich jede Theorie der Theorienreduktion die Möglichkeit der Multiplen Reduzierbarkeit erlauben müsse, – eben weil diese nicht die Ausnahme, sondern eher die Regel sei. Bickle greift dabei auf Hookers (1981) Theorie der *token-token reduction* zurück, d.h. auf die Annahme, dass ein mentaler Zustand zu einem bestimmten neurobiologischen Zustand korrespondiert. Es sei T_R die reduzierte Theorie und T_B die reduzierende Theorie[70] (Beispiel:

[70]a reducing theory is a more general theory that includes a reduced theory as a special case, while a reduced theory is a specific case of a more general theory obtained by imposing

Einsteins Mechanik (SRT) ist die reduzierende Theorie, Newtons Mechanik als Spezialfall der Einsteinschen ist die reduzierte Theorie (Stegmüller: Die klassische Partikelmechanik ist für den Grenzfall von Geschwindigkeiten, die im Verhältnis zur Lichtgeschwindigkeit klein sind, auf die relativistische Mechanik reduzierbar, dabei ist die statistische Mechanik die allgemeinere, die Thermodynamik die reduzierte Theorie). Im Folgenden soll der Index B anzeigen, dass es sich um die "Basistheorie" handelt, und der Index R, dass es sich um die reduzierte Theorie handelt. Mit T_B wird nun eine zu T_R analoge Theorie T_R^* formuliert, die unter mit C_R bezeichneten Bedingungen die syntaktische Struktur der fundamentalen Gesetze von T_R widerspiegelt; die C_R sind insbesondere Grenzbedingungen oder eingrenzende Annahmen. T_R^* ist dann eine *deduktiv valide* Folgerung von $T_B \wedge C_R$ (wobei $\wedge$ für die Konjunktion "und' steht), und eine analoge Relation A_R zwischen T_R und T_R^* gilt. Während C_R sehr komplex sein kann besteht die Beziehung zwischen T_B und T_R^* aus einer sehr direkten logische Ableitung derart, dass T_B die Theorie T_R^* unmittelbar erklärt und die über die Relation A_R die Beziehung zwischen T_R und T_R^* erklärt. T_B erklärt somit indirekt T_R.

Als Beispiel werde wieder die Reduktion der Thermodynamik auf die Statistische Mechanik betrachtet. C_R bezeichne das thermodynamische Limit. T_R^* sei die Menge von Theoremen aus einem beschränkten Bereich T_B der Statistischen Physik, die die syntaktische Struktur der fundamentalen Gesetze der Gleichgewichtsthermodynamik T_R imitieren relativ zu der vorher festgelegten Analogie A_R. In klassischen Theorie der Reduktion ist T_R die reduzierte Theorie, in Hookers Theorie ist T_R^* die reduzierte Theorie. Der große Vorteil von Hookers Theorie besteht darin, nicht den Status von Brückengsetzen erklären zu müssen, – wobei aber u.a. noch geklärt werden muß, ob Brückengesetze eigentlich Definitionen oder analytische Wahrheiten sind, oder kontingente Identitäten, oder ob sie einfach nur nützliche Konventionen darstellen (Bickle (1992b), p. 55). Damit wird auch Fodors Argument der "wild disjunktiven" Brückengesetze) ausgeräumt, das heißt, so Bickle, es wird eine Reduktion der Psychologie auf "lower-level" Wissenschaften wie die Neurobiologie möglich, die behauptete Multiple Realisierbarkeit der Psychologie stellt also keine Hürde für ihre Reduzierbarkeit auf die Neurobiologie mehr dar, – abgesehen davon, dass die in Abschnitt 2.4.3.3 beschriebenen empirischen Befunde die Multiple Realisierbarkeitsowieso als ein fragwürdiges Konzept erscheinen lassen.

2.4.3.7 Der Hooker-Churchland-Ansatz Bereits P.M. Churchland (1985), und Richardson (1982) hat die Charakterisierung der multiplen Realisierbar-

additional constraints or assumptions. Alternative (wiki): Philosophers use "theory reduction" as a term of art to denote the scientific practice whereby a more basic theory expresses or otherwise captures the facts and principles described by a less basic theory. The reducing theory thus preserves the ontology of the reduced theory, at least in ideal cases.

keit über den Begriff der Bikonditionalität als viel zu eng kritisiert: gemäß der Nagelschen Theorie der Reduktion wird eine ältere Theorie durch eine neuere, umfassendere Theorie dann reduziert, wenn die neue Theorie zusammen mit geeigneten Korrespondenzregeln die Prinzipien der alten Theorie logisch impliziert. Diese Regeln sind die Brückengesetze. In der Churchlandschen Sprechweise verbinden die Brückengesetze die verschiedenen Ontologien der beiden Theorien: Temperatur auf der einen Seite und kinetische Energie auf der anderen Seite, d.h. Temperatur sei *identisch* mit der mittleren kinetischen Energie. In festen Körpern können Moleküle allenfalls vibrieren, so dass Temperatur nur auf oszillatorische Prozesse zurückgeführt werden kann, und im Vakuum gibt es gar keine Moleküle, so dass Temperatur auf die Vakuumstrahlung bezogen werden muß, und die wirklichen (also nicht die in der Theorie angenommenen) Gase gehorchen nicht der Gleichung $PV = \mu RT$ der klassischen Thermodynamik. Wegen der in Keplers Gesetzen nicht explizit genannten gravitationalen Wechsewirkungen zwischen den verschiedenen Planetenbewegen und einiger anderer Effekte bewegen sich die Planeten des Sonnensystems nicht streng auf den in Keplers Gesetzen postulierten Ellipsenbahnen, und die Beschleunigung eines fallenden Körpers ist gemäß dem Galileischen Fallgesetz nur im Vakuum eine Konstante, außerhalb des Vakuums kann der Luftwiderstand einen Einfluß haben. Wäre also, wie Nagel und mit ihm Fodor annehmen, die Reduktion eine Deduktion der Form $p \rightarrow q$, d.h. wenn die Aussage $p$gilt, dann auch die Aussage q, so würde nach der Schlußregel des *modus tollens* $\neg q \rightarrow \neg p$ folgen, dass die reduzierenden Theorien (Thermodynamik, Newtons Mechanik) ebenfalls falsch sind: Folgt aus der Temperatur T ein bestimmter Wert der kinetischen Energie E_{kin}, so folgt aus einer von E_{kin} abweichenden Energie, dass auch die Temperatur *nicht* durch T gegeben ist. Dies bedeutet, dass die Reduktion als syntaktische Beziehung zwischen den logischen Strukturen der reduzierten Theorie T_0 und der reduzierenden Theorie T_N aufgefasst werden kann. Was also den H-C-Ansatz vom Brücken-Gesetz-Ansatz für die Reduktion unterscheidet ist die Zielsetzung der Reduktion: die reduzierte Theorie ist T_0, und die reduzierende Theorie Z_N dient als Prämisse. Im Unterschied dazu wird im H-C-Ansatz die analoge Struktur I_N hergeleitet, die bereits im Vokabular von T_N spezifiert wurde. I_N ist ein approximativ "äquipotent isomorphes Bild von T_0", – es ist approximativ, weil die meisten Reduktionen Korrekturen der reduzierten Theorie enthalten. Also ist es nie die reduzierte Theorie T_0, die deduziert wird, sondern T_0 ist das Ziel einer *komplexen Mimikry*. I_N wird bereits in einem Teil der Sprache von T_N spezifiziert, also gibt es keine inter-theoretischen (cross-theoretic) Aussagen, die die nicht-trivialen Ableitungen beeinflussen.

Der H-C-Ansatz vermeidet die Spezifikation des logischen Status des kreuztheoretischen Identitätsaussagen, der für den auf Brückengesetzen beruhenden Ansatz schwierig wird, falls eine Reduktion bedeutende Korrekturen der reduzierten Theorie enthält. Denn kreuz-theoretische Identitätsaussagen spielen

in der Herleitung der Analogstruktur I_B keine Rolle.

Anmerkung: Davidsons These, dass es keine psychophysischen und allgemein keine psychologischen Gesetze gebe, impliziert, dass sich die Psychologie nicht auf basalere Wissenschaften wie die Neurobiologie reduzieren lässt. Dieses Argument setzt stillschweigend voraus, dass der Begriff der Reduktion die Notwendigkeit von Brückengesetzen impliziert, – und diese wiederum basieren auf einem bestimmten Begriff von Theorie (es geht ja um intertheoretische Reduktion). Die Nagelsche Theorie der Reduktion, die bei Diskussionen um die Reduzierbarkeit unterstellt wird, beruht u.a. auf einem bestimmten Konzept von Theorie: damit z.B. eine Sammlung psychologischer Gesetze zu einer psychologischen Theorie zusammengefasst werden können, müssen die Gesetze logisch konsistent, d.h. miteinander logisch verträglich sein. Über diese Kriterien hatten sich die Philosophen des Logischen Empirismus Gedanken gemacht und ein Konzept von Theorie entwickelt, für das der Ausdruck *set-of-sentences* geprägt wurde (der englische Ausdruck soll hier beibehalten werden, weil der gelegentlich verwendete Ausdruck *Satzmengentheorie* mehrdeutig ist, denn er wird mit Begriffen aus der mathematischen Mengenlehre assoziiert). In der set-of-sentences-Theorie von Theorien geht es um die Frage der Repräsentation von Wissen. Diesem Konzept von Theorie entsprechend sind Aussagen der Psychologie einerseits und den letzlich physikbasierten Wissenschaften andererseits von unterschiedlichem linguistischen Typ: es gibt Nichtgesetze (nonlaws) einerseits und Gesetze (laws) andererseits, – und dieser Unterschied "versenkt" die Möglichkeit der Reduktion. Denn Reduktion erfordert, dass Gesetze der einen Wissenschaft auf Gesetze einer anderen Wissenschaft abgebildet werden. Will man gleichwohl eine Reduktion ermöglichen muß der set-of-sentences-Ansatz vermieden und durch einen anderen Ansatz ersetzt werden.

Ein anderer Aspekt *des Neuen Reduktionismus* ist der Sachverhalt, dass Anhänger dieser Richtung entschiedene Gegner der Vorstellungen der logischen Empiristen sind, wobei sie aber die Auffassung der logischen Empiristen, derzufolge Theorien Mengen von Sätzen (im Sinne von Theoremen) sind, nicht explizit kritisieren (Bickle (1992a), Mental Anomaly). Diese Konzeption von Theorien ist aber die Kernannahme des logischen Empirismus. s. p. 226 - 227: dieser Ansatz sei, so Bickle, "seriously flawed"[71]. Im sets-of-sentences-Ansatz wird angenommen, dass psychische Prozesse gewissermaßen linguistisch repräsentiert werden, und dass "Verstehen" darin besteht, logische Implikationen aus Teilmengen solche Sätze zu bestimmen. Mit diesem Ansatz sind aber starke Annahmen über die Speicherung und den effizienten Abruf von Informationen verbunden. Es müssen riesige Mengen von "Sätzen" und

[71] "First, already a growing consensus among some new reductionists that the sets-of-sentences account of scientific theories specifically, of knowledge representation generally, is seriously flawed. (Paul Churchland's recent essays are a key result; see especially Churchland 1988, 1989)", Bickle (1992), p. 226

deren Verallgemeinerungen gespeichert werden, um kontinuierlich Erlerntes abzuspeichern, dazu kommt das Problem des Abrufs solcher Sätze. Je länger die Listen von Sätzen sind, je langsamer wird der Abruf, im Widerspruch zu dem Befund, dass Menschen mit wachsender Erfahrung in ihren Reaktionen immer schneller und genauer werden (Nisbett & Wilson (1977)), – s.a. auch das Zitat von PM Churchland zur Kritik an Set-of-sentences-Theorie.

Dem H-C-Ansatz zufolge besteht eine intertheoretische Reduktion in einem Beweis, dass "eine mehr umfassendere Theorie T_N erklärende und prädiktive Komponenten enthält, die der reduzierten Theorie T_0 entsprechen" (Bickle (1996), p. 222). Wie in der ursprünglichen Theorie der Reduktion bleibt die syntaktische Beziehung zwischen den logischen Strukturen der beiden Theorien erhalten: die reduzierende Theorie T_N dient als Prämisse und die reduzierte Theorie T_0 ist das Ziel der Deduktion – diese Deduktion erfordert Brückengesetze. Wie oben ausgeführt wurde wird beim H-C-Ansatz eine bereits in T_N enthaltene Analogstruktur I_N deduziert, die ein "approximativ äquipotentes isomorphes Bild" (Bickle) von T_0 ist; das Bild ist approximativ, weil Reduktionen im Allgemeinen Korrekturen der reduzierten Theorie implizieren.

Nach Davidson existieren keine psychophysischen Gesetze und damit auch keine Brückengesetze, so dass das Mentale nicht auf das Neurobiologische reduziert werden könne. Aber Brückengesetze und damit auch psychophysische Gesetze werden Dem H-C-Ansatz zufolge gar nicht für eine Reduktion benötigt. Gleichwohl kann Davidson nun fragen, was für eine Art von Theorie es denn sei, die keine Gesetze enthalte: – eine derartige Theorie sei eben gar keine Theorie, oder, anders gesagt, die Theorie besteht aus der Aussage, dass es keine Theorie gibt. Eine Antwort auf dieses Argument besteht einfach darin, Davidsons Theorie der Nichtexistenz psychophysischer Gesetze in Frage zu stellen. Bickle weist darauf hin, dass wissenschaftlichen Theorien üblicherweise gar nicht so konzipiert sind, wie von den Nicht-Reduktionisten angenommen wird, – diese Denker legen ihrer Kritik einfach das Modell der Logischen Empiristen, d.h. der Theorie der *Sets-of-sentences*-Präsentationen von Wissen, zugrunde. Damit ist die von den logischen Empiristen des Wiener Kreises entwickelte Konzeption des Begriffs einer wissenschaftlichen Theorie gemeint: Es werden zunächst grundlegende Annahmen bzw. Axiome festgelegt derart, dass sich die Aussagen (die "sentences") logisch aus aus den Annahmen und Axiomen ergeben. Damit soll unter anderem die logische Konsistenz der Theorie gesichert sein. Diese Konzeption wissenschaftlicher Theorien ist bekanntlich nicht unkritisiert geblieben, weil ihre Orientierung an formalen Theorien dem Charakter empirischer Theorien nicht gerecht wird. Empirische Theorien ensprechen eher dem Kuhnschen Begriff des wissenschaftlichen Paradigmas. Allein deswegen schon hat Davidsons Theorie keinen Allgemeinheitsanspruch und scheidet als Gegenargument gegen den Reduktivismus aus.

2.4.4 Zusammenfassende Kritik

Bevor auf spezielle Kritikpunkte eingegangen wird, soll ein kurzer Überblick gegeben werden. Zuerst wird eine globale Kritik des MR-Konzept gegeben:
(i) Bickles Drei Fehler, dann
(ii) die Abstraktheit, oder besser die Abgehobenheit philosophischer Argumente: MR wird einfach postuliert. So stellte zum Beispiel Zangwill (1992) fest, dass MR über Species empirisch nie nachgewiesen wurde. Woher Philosophen die Sicherheit nehmen, mit der sie behaupten, dass sich Bewußtsein siliziumbasiert erzeugen lässt, ist nicht bekannt. Ohne weitere Kenntnisse wäre es doch sinnvoll, zunächst die biochemischen Prozesse zu untersuchen, die tatsächlich involviert sind, wenn Bewußtsein entsteht.
(iii) Die Schlichtheit der MR-These: sie vernachlässigt die stochastischen Komponenten. Spiking-Modelle sind wichtig, aber unvollständig, weshalb der Hinweis auf die Biochemie wichtig ist.

2.4.4.1 Fodors Fehler Philosophen oder Wissenschaftler, die der Ansicht sind, mentale Ereignisse könnten nicht auf physische Prozesse zurückgeführt werden, verweisen oft auf nicht akzeptierbare Implikationen der These, mentale Ereignisse oder Prozesse seien auf physische Prozesse zurückführbar. Bickle (1996) hält diese Implikationen aber für falsch; er betrachtet insbesondere drei derartige Fehlschlüsse:

Fehler 1: Der Reduktionismus impliziert Einschränkungen bezüglich der Wahl der möglichen Theorien in den Special Sciences.
Fehler 2: Die Reduktion einer Theorie T_n auf eine andere Theorie T_B bewirkt das Verschwinden der reduzierten Theorie T_n.
Fehler 3: Die Ontologie der Reduktion impliziert die Nutzlosigkeit der Psychologie.

Die dem Fehler 1 unterliegende Argumentation ist eigenartig. Denn angenommen, die Psychologie (oder eine andere "special science") sei tatsächlich auf die Neurobiologie reduzierbar. Dann haben mentale Ereignisse eine neurobiologische Basis. Das heißt aber nur, dass Hypothesen, die grundsätzlich nicht mit einer biologischen Basis kompatibel sind, eine eher kleine a priori-Wahrsheinlichkeit zugewiesen bekommen – warum das ein Nachteil sein soll, ist unklar. Denn keine Theorie ist kompatibel mit jeder anderen Theorie, d.h. wenn man einer bestimmten Theorie zuneigt, verringert sich die Zuneigung zu den anderen Theorien; dieser Befund ist eine Trivialität. Er gilt auch für Fodors Theorie der Nichtreduzierbarkeit der higher-level-Wissenschaft Psychologie, – akzeptiert man diesen Ansatz, so reduziert diese Entscheidung die Wahrscheinlichkeit, die Psychologie als reduzierbare Wissenschaft zu betrachten. Das Argument, mentale Prozesse seien auf verschiedene Weise realisierbar ist aber nicht zwingend und vermutlich einfach falsch, es gibt also keinen

Grund, der Fodorschen Argumention zu folgen.

Fodor geht, wie gezeigt wurde, von einem durch Bikonditionalitäten gekennzeichneten Reduktionsbegriff aus. Bickle (1996) verweist darauf, dass die Thermodynamik sich im Rahmen der Phänomenologie der Wärme entwickelt habe, die Entwicklung der statistischen Physik aber in einem anderen Rahmen stattgefunden habe und man erst im Anschluß dieser Entwicklungen die Reduzierbarkeit der einen auf die andere Theorie gefunden habe, wobei die Multiple Reduzierbarkeit (MR) thermodynamischer Größen keine Rolle gespielt habe. Für Fodor ist aber die MR ein wesentlicher Punkt in der Argumentation der Nichtreduzierbarkeit mentaler Ereignisse. In der Physik ist aber die MR eher die Norm als die Ausnahme und trotzdem ist die Thermodynamik auf die statistische Physik reduzierbar. Bickle (1996, p. 63) weist dementprechend darauf hin, dass sich wissenschaftliche Theorien nicht nach Maßgabe eventueller Reduzierbarkeit entwickelt haben; insofern könne von einem einschränkenden Effekt der Reduzierbarkeit keine Rede sein.

Die den Fehler 2 ausmachende These geht ebenfallls auf Fodor (1975) zurück. Er postuliert, dass die Psychologie aus der Wissenschaft verschwände, könne sie auf die Neurobiologie zurückgeführt werden. Fodors Befürchtung beruht, wie es scheint, wiederum auf seinem sehr speziellen Begriff von Reduktion. Wäre eine Reduktion in der von Fodor angenommenen Art möglich, so könnte man vielleicht tatsächlich fragen, ob die Psychologie die Eigenständigkeit hat, von der man heute noch ausgeht. Am Beispiel der Thermodynamik kann man aber feststellen, dass sie seit ihrer Reduktion auf die statistische Physik keineswegs überflüssig geworden ist. Würde die Thermodynamik überflüssig, so würde man zum Beispiel den Carnotschen Kreisprozess im Rahmen der statistischen Physik behandeln, was für die Konstruktion von Wärmepumpen mit größerer Umständlichkeit verbunden wäre, – hätte man die Thermodynamik nicht schon zur Verfügung, würde man sie vermutlich erfinden. Eine analoge Betrachtung ergibt sich für die Chemie: im Grundlagenbereich geht die Chemie in die Physik über, aber deswegen wird die Chemie nicht notwendig von ihren Grundlagen in der Physik aus betrieben. In ähnlicher Weise wird die Psychologie, auch wenn psychologische Prozesse durch neurobiologische Prozesse erklärt werden können, ihre Theorien nicht stets und grundsätzlich aus der Neurobiologie herleiten. Die Reduktion der Psychologie auf die Neurobiologie dient im Allgemeinen einem tieferen Verständnis psychologischer Prozesse, was zu besserem makropsychologischen Verständnis führen kann. Ein Beispiel sind Wahrnehmungstäuschungen oder Gestalteffekte nicht nur in der Wahrnehmung, wo man versucht, für derartige Phänomene die neurobiologischen Mechanismen zu bestimmen, die diese Phänomene erzeugen. Dabei werden für praktische Fragestellungen, etwa in der Verkehrspsychologie, die *psychologischen* Gesetze relevant sein und die neurobiologischen Prozesse werden eine eher sekundäre Rolle spielen; Bickle (1996, pp. 65 - 67) liefert eine Reihe weiterer Beispiele.

Die dem Fehler 3 unterliegende Annahme wird bereits durch die vorangegangenen Betrachtungen widerlegt. Es ist aber nützlich, sich die immanente Widersprüchlichkeit dieser These klarzumachen. Denn es wird ja argumentiert, dass jede psychische Dynamik eben nur auf dem Niveau ("level") der Psychologie gefunden und beschrieben werden könne, woraus wiederum die Unmöglichkeit bzw. Sinnlosigkeit einer Reduktion der Psychologie auf die Neurobiologie folge. Das Gegenargument ergibt sich aus der Analogie zur Beziehung zwischen Thermodynamik und statistischer Physik. Kaum jemals wird ein Mensch, der sich seinen Morgenkaffee zubereitet, versuchen, die Temperatur des Wassers durch die Bestimmung der kinetischen Energie von Wassermolekülen einer "hinreichend großen" Stichprobe von H_2O-Molekülen abzuschätzen, die Beobachtung von siedendem Wasser ist völlig hinreichend für die Zwecke der Kaffeezubereitung, und ebenso genügt ein Quecksilberthermometer, will man die Körpertemperatur bei einem grippalen Infekt bestimmmen. In analoger Weise ergibt sich etwa das Verständis von Fehlern, die von Verkehrsteilnehmern im Straßenverkehr begangen werden, aus dem phänomenologischen Befunden bezüglich Wahrnehmungs- und Aufmerksamkeitstäuschungen, bei denen die Kenntnis der korrespondierenden neurobiologischen Ursachen das Verständnis vertiefen kann (und eventuell entsprechende Maßnahmen begründen kann), die aber nicht explizit betrachtet werden müssen, um Fehlverhalten zumindest probabilistisch voraussagen zu können. Für die Abschätzung der Validität von Zeugenaussagen in Kriminalprozessen gelten ähnliche Überlegungen: wenn neurobiologische Grundlagen der psychologischen Prozesse diskutiert werden, geht es in erster Linie um Erklärungen und sehr viel weniger um Voraussagen von Verhalten. Eine radikale Ablehnung jeglicher Form von Reduktion auf neurobiologische Grundlagen des Verhaltens macht deshalb keinen Sinn, das Bestehen auf quasi-dualistischen Theorien führt dagegen mit erhöhter Wahrscheinlichkeit in die Sterilität hermeneutischer Beliebigkeiten. Es ist, als wolle man die Funktionsweise eines Radios "verstehen", indem man notiert, welche "Antwort" das Gerät gibt, wenn man den Knopf für die Wahl von Sendern dreht zu verschiedenen Zeitpunkten und geographischen Orten dreht: bei bestimmten Positionen zu bestimmten Zeitpunkten und an bestimmten Orten hört man eine bestimmte Sprache, bei einer anderen Position auf der Senderskala hört man Musik, an wieder anderen Positionen hört man nur ein Prasseln oder Rauschen, und das "Design", also der Versuchsplan, für solche Messungen erlaubt dann, Unterschiede von Mittelwerten, Wechselwirkungseffekte etc zu bestimmen. Schaut man in das Innere des Radios, sieht man entweder Platinen mit aufgedruckten Mustern von metallischen Verbindungen, oder, wenn es sich um ein älteres Modell handelt, verstaubte Glaskolben und Drähte. Man schließt auf die multiple Realisierbarkeit, wenn bei bestimmten Skalenpositionen nur Rauschen zu hören ist, bei anderen Positionen Sprachen, und bei wieder anderen Musik aus dem Lautsprecher kommt und stellt eine Nichtreduzierbarkeit der akustischen Phänomene auf das elektronische Geschehen fest. Professor Alva Noë, von dem insbesondere bei der Diskussion der Fra-

ge nach der Existenz eines freien Willens noch die Rede sein wird, und seine philosopischen Anhänger triumphieren: der ganze elektronische Klimbim im Gehäuse des Radios hat nichts mit der wunderbaren, lebendigen Vielfalt des Geschehens am Lautsprecher zu tun!

2.4.4.2 Allgemeines zur MR-These: Überhaupt ist die MR-These in ihrer suggestiv wirkenden Einfachheit außerordentlich radikal. Es wird postuliert, dass für ein gegebenes mentales Ereignis physische Prozesse $\phi_1, \ldots, \phi_p, \ldots$ existieren können derart, dass $M = M(\phi_1) = \cdots = M(\phi_p) = \cdots$, obwohl $\phi_j \neq \phi_k$ für $j \neq k$, – M ist multipel realisierbar. Die Aussagen der Philosophen, die diese These ohne weitere Spezifizierung teilen, ergeben sich intuitiv aus sehr allgemeinen Annahmen, etwa aus der These, dass Kognitionen funktionalistisch charakterisiert werden können und *deswegen* der Bezug auf eine bestimmte physische Basis der Kognitionen entweder überflüssig oder gar nicht möglich sei. Empirisch arbeitende Wissenschaftler werden in der MR-These eher eine Hypothese mit vernachlässigbarer Wahrscheinlichkeit, wahr zu sein sehen. Die von Putnam immer wieder als Beispiel herangezogenen C-Fasern für die Schmerzerfahrung sind empirischen Befunden entprechend für Mensch und Tier im Wesentlichen identisch (Larsson et al (2022)), was vermuten läßt, dass die Schmerzerfahrungen von Mensch und Tier sehr ähnlich sind, – es sei denn, es gibt Unterschiede im Prozess der Bewußtwerdung. Die zahlreichen Untersuchungen Hubel und Wiesels zum Aufbau und der Funktionsweise des visuellen Systems der Katze und der großen Ähnlichkeit dieses Systems mit den visuellen Systemen der Primaten (einschließlich des Menschen) haben durchaus zu Einsichten über Wahrnehmungsprozesse geführt, wobei allerdings die anatomischen Befunde aus Sicht der Anhänger der MR-These als überflüssig und belanglos erklärt werden - in sehr radikaler Weise von dem oben bereits genannten Philosophen Alva Noë, der mit der These, menschliche Kognitionen hätten gar nichts mit dem Gehirn zu tun, seinen Bekanntheitsgrad deutlich erhöht hat. Nach Putnam, Fodor und anderen soll die MR-These für verschiedene Species gelten, vom homo sapiens bis zur Eidechse und vielleicht auch bis zum Regenwurm soll zum Beispiel Schmerz (= M) immer ein und dasselbe mentale Ereignis sein, das durch die entsprechend verschiedenen Hirnpozesse ϕ_k multipel realisiert wird. Warum das so sein soll, wird nicht erklärt oder begründet, die These wird schlicht behauptet. Denkbar ist jedenfalls, dass sich die durch die ϕ_k erzeugten mentalen Ereignisse durchaus voneinander unterscheiden (am Beispiel Schmerz sind die ϕ_k Prozesse, die den Organismus dazu bringen sollen, mit Dringlichkeit einer für den Organismus schädlichen Situation zu entkommen, – "Schmerz" könnte diese Dringlichkeit repräsentieren. Aber natürlich ergibt sich die Frage, wie man denn feststellen kann, ob eine Empfindung, für die wir in Bezug auf den Menschen einen bestimmten Ausdruck haben, über die Species hinweg dieselbe Qualität hat. Wenn wir keine Möglichkeit haben, die Identität dieser Qualität

festzustellen, verliert die reine Behauptung dieser Identität ihren Sinn, da ein rein philosophischer Beweis nicht geliefert wird und aller Wahrscheinlichkeit nach auch gar nicht konstruiert werden kann. Nicht auszuschließen ist doch, dass 'Schmerz' nur eine von sehr vielen möglichen Reaktionen auf die Umwelt ist. Berühmt ist Nagels (1974) Diskussion der Frage, wie eine Fledermaus ihre Umgebung wahrnimmt, da ja die Konstruktion der mentalen Repräsentation der Umgebung in erster Linie auf akustischen Signalen beruht.

Visuelle Wahrnehmung und MR: Schon sehr "einfache" Lebewesen erhöhen ihre Überlebenswahrscheinlichkeiten durch Registrierung optischer Signale. Die Evolution hat das Organ 'Auge' auf viele verschiedene Weisen etwa als Facettenauge bei Insekten und Crustacea oder als Kameraauge bei Wirbeltieren sich entwickeln lassen. Das Auge des Octopus ist ein Kameraauge, das in vielerlei Hinsicht mit dem Auge des Menschen vergleichbar ist. Es enhält polarisierungsempfindliche Zellen (wie sie auch die Facettenaugen von Insekten besitzen), das die Richtung der Polarisierung des einfallenden Lichts zu bestimmen gestattet und somit für die Navigation des Tieres von Vorteil ist. Das Gehirn eine Octopus ist ähnlich organisiert wie das Gehirn von Wirbeltieren, die Tiere können Probleme lösen und sozial interagieren, so dass Anlass besteht, dass sie eine Form von Bewußtsein haben. Das Postulat, dass die vermutlich existierenden mentalen Prozesse des Octopus unabhängig von seinem Gehirn sind liefert keinerlei Einsicht in die Natur der Beziehung zwischen Gehirn und Mentalität. Dass Behauptungen wie die der Multiplen Realisierbarkeit unerwünschte Konsequenzen haben können illustriert ein Beispiel aus der Medizin: so wurde etwa die Epilepsie zuerst als supranatürliches Phänomen angesehen und erst ab Ende des 19ten Jahrhunderts als Abnormalität organischer Natur erkannt. Epilepsie wurde als supranatürliches Phänomen hingenommen, weil sie *per Definition* als nicht medizinisch behandelbar galt. Erst als sie aber als physisches Phänomen erkannt wurde, konnte man sich auf die Suche nach medizinischer Hilfe machen.

2.4.4.3 Gartenhecken, Granularität, Stochastizität Man muß das Prinzip der Multiplen Realisierbarkeit (MR) in seiner radikalen Form, der zufolge mentale Ereignisse unabhängig von der Hirntätigkeit sind, nicht akzeptieren, schon weil eine Unzahl medizinischer Befunde gegen ein solches Prinzip spricht. Man kann argumentieren, dass die MR-These ein Resultat von zu abstrakter, um nicht zu sagen abgehobener Argumentation ist, bei der von einem radikalen Identitätsbegriff leibnizscher Prägung[72] einerseits und dem Begriff der Turingmaschine andererseits ausgehend auf die multiple Realisierbarkeit geschlossen wird. Man muß aber einräumen, dass hochkomplexe Organe wie das Gehirn unter identischen äußeren Bedingungen zu verschiedenen Zeitpunk-

[72] Ein Gegenstand x ist genau dann mit einem Gegenstand y identisch, wenn alle Prädikate (Eigenschaften) F, die x zukommen, auch y zukommen und umgekehrt. Leibniz, G. W., Philosophical Papers and Letters, in Loemker 1969.

ten nicht in exakt derselben Weise aktiv sind, aber – um ein einfaches Beispiel zu nehmen – die Wahrnehmung von Objekten unter identischen Bedingungen nicht variiert. Es ist argumentiert worden, dass allein schon dieser Sachverhalt eine Rechtfertigung des MR-Prinzips sei (Beckermann (2008)). Dieser Folgerung muß man nicht folgen, aber es ist klar, dass die mentale Invarianz zum Beispiel gegenüber stochastischen Fluktuationen der neuralen Aktivität einer Erklärung bedarf. Tatsächlich ist die Stochastizität bei Modellierungen neuronaler Aktivität selbstverständlich (Dayan & Abbot (2001), Gerstner & Kistler (2002), Murray (1989)), Grossberg (2021). Die Identitätstheorien der fünfziger Jahre sind noch verhältnismäßig schlicht, es ist illustrativer, von den wesentlich komplexeren Vorstellung der neueren Hirnforschung auszugehen, wie sie z.B. Singer (2007) vorgestellt hat: hier geht es um die Synchronisierung von über das Gehirn verteilter Aktivität, wobei aber die intrinsische Stochastizität neuronaler Aktivität nicht explizit mitdiskutiert wurde. Mit Bezug auf die MR-These und den genannten strengen Identitätsbegriff sollte aber zur Kenntnis genommen werden.

Shapiro & Polger (2012) führen dazu den Begriff der "kompositionellen Variation"[73] von Hirnzuständen ein. Darunter verstehen sie eine Variation des Hirnzustands, die mit einem bestimmten mentalen Zustand kompatibel ist. Zur Illustration verweisen Shapiro & Polger auf die in englischen Gärten oft zu findenden Buchsbaumhecken, die so beschnitten werden, dass eine bestimmte Form – etwa ein Hahn oder ein Pfau – entsteht. Diese Formen sind i. A. einander so ähnlich, dass man von einer Hahnen- bzw. Pfauenhecke sprechen könne, d.h. die mentalen Zustände, die der Wahrnehmung einer solchen Hecke entsprechen, sind in dem Sinne identisch, dass sie der Kategorie 'Hahn' oder 'Pfau' entsprechen. Tritt man allerdings hinreichend nahe an eine solche Hecke heran, so sieht man die einzelnen Äste der Hecke, und diese "Nahaufnahmen" von zwei Hecken sind zweifelsfrei nicht identisch im Sinn der strengen Definition von 'Identität'. Die Autoren sprechen in diesem Zusammenhang von der Granularität von Phänomenen, und diskutieren unter Bezug auf die Granularität, was MR über verschiedene Species hinweg bedeutet. Damit ist die These gemeint, dass ein bestimmtes mentales Ereignis M – z.B. eine Wahrnehmung, oder eine Furchtreaktion – für verschiedene Species identisch ist, obwohl sich die neuronalen Basen der Species voneinander unterscheiden. Falls M für verschiedene Species wirklich identisch sein kann. Eine alternative Annahme wäre, dass bei allen Unterschieden der phylogenetischen Entwicklung die zu bewußten Wahrnehmungen korrespondierenden molekuaren Mechanismen einander sehr ähnlich sind.

Couch (2004) betrachtet Primatenaugen im Vergleich zu Octopusaugen. Diese Augentypen unterscheiden sich hinsichtlich der Pigmente in den Photorezeptoren, hinsichtlich ihrer Retinae und der Mechanismen, die die Fokussierung von Licht besorgen. Der MR-These entsprechend handelt es sich um

[73] compositional variation

multiple Realisierungen eines allgemeinen Augenprinzips. Andererseits erzeugen die beiden Augentypen für gegebenen Input unterschiedliche outputs, – sie erzeugen nicht dieselbe Wahrnehmung. Die Wahrnehmungen mögen *ähnlich* sein – aber sie ist nicht *identisch*, wie vom Putnamschen MR-Prinzip verlangt. Psychologen und Neurowissenschaftler schließen i.A. von neuronalen Unterschieden auf unterschiedliche Funktionen (Reaktionen). In der wissenschaftlichen Praxis sei es so, dass Unterschiede in der neuronalen Repräsentation auf funktionale Unterschiede verweisen; Couch betont, dass die Entstehung psychologischer Zustände einerseits und neuronaler Zustände andererseits eine empirische Frage ist, und sowohl Couch wie Shapiro (2000) stellen fest, dass Philosophen keine genauen Analysen durchführen, sondern sich einfach auf die Alltagspsychologie beziehen (Bickle (2020), Abschnitt 2.4). Northoff und Heinzel (2006) argumentieren, dass sich die Variationen in der neuronalen Aktivität praktisch nicht auf das korrespondierende mentale Ereignis auswirken. Allerdings stützt dieser Befund nicht die MR-These, sondern eher die Hypothese der glättende Rolle der biochemischen Transformation von neuronaler Aktivität in bewußtes Erleben.

William Bechtel and Robert McCauley (1999) entwickelten eine "heuristic mind-brain identity theory (HIT)" im Unterschied zur MR-Theorie: Wissenschaft untersucht typischerweise Hypothesen, die im Laufe empirischer Untersuchungen aufgestellt werden und die die Richtung für die folgenden Untersuchungen angeben. Dass nicht nur die klassischen neuronalen Aktivitäten beobachtet werden müssen belegen Studien zu molekularen Prozessen (Bickle 2006)), die seit langer Phylogenese auf uns gekommen sind. Womöglich sind die basalen physiologischen Prozesse, die Bewußtsein erzeugen, bei allen Wesen identisch[74], – was aber noch nicht bedeutet, dass die mentalen Ereignisse identisch sind, denn sie hängen, wie beim Auge des Octopus, vom spezifischen Input ab. Mit der generellen MR-These sind derlei Befunde nicht kompatibel.

2.4.4.4 PM Churchlands Abrechnung Churchland betrachtet zunächst die *Sieben Thesen*, die den komputationalen Funktionalismus charakterisieren:

1. Was allen kognitiven Kreaturen gemein ist, ist nicht, dass sie die gleichen komputationalen Mechanismen (= hardware) haben, sondern dass sie im Wesentlichen dieselben Input-output-Funktionen berechnen (mit analogen Zustandsübergängen)

2. Der Funktionalismus erklärt als Ziel der Kognitionswissenschaft die Berechnung der *abstrakten Funktion*, die uns als kognitive Wesen definiert,

3. Die zentrale Aufgabe der Künstlichen Intelligenz (KI) ist es, neue *physikalische Realisierungen* zu finden, die die salienten Teile dieser Funktion bestimmen,

[74] Fliegen in den Kopf geschaut https://www.uni-muenster.de/news/view.php?cmdid=9990

4. Die *Folk Psychology*[75] ist die "normale" Psychologie, die wir alle täglich anwenden, um unsere kognitiven Aktivitäten zu strukturieren,

5. Die Reduktion der Psychologie auf die Neurophysiologie ist aus zwei Gründen unmöglich, (i) die relevante Funktion ist auf unendlich viele Weisen berechenbar, nicht nur in der Weise, die man bei Menschen findet,

6. Die empirische Forschung in die Mikrostruktur menschlicher und tierischer Hirne ist zwar legitim, ist aber eine schlechte Forschungsstrategie, wenn die globale Funktion identifiziert werden soll, die sich auf eher molarem Verhaltnsniveau äußert.

7. Die Punkte (5) und (6) verpflichten uns, die methodologische Autonomie der Kognitiven Pschologie zu "respektieren und zu verteidigen", – in Abgrenzung zu den "low-level" Wissenschaften Anatomie, Physiologie, und Biochemie. Die Kognitionswissenschaften suchen nach ihren eigenen Gesetzen; dies ist Fodors (1974) *Argument der methodologischen Autonomie* der Psychologie.

Churchland räumt ein, dass die sieben Punkte zwar eine sehr einflußreiche philosophische Position repräsentieren, – aber alle sieben Punkte seien erstaunlicherweise einfach falsch.

Die wesentlichen Argumente gegen den funktionalen Komputationalismus ergeben sich für Churchland zunächst aus einem Vergleich neurobiologischer Organe mit dem Computer. Letztere wurden immer schneller (desk-top über 100^9 Hz), immer größere Speicher (10^{19} Bytes, die Signalübertragung findet nahezu mit Lichtgeschwindigkeit statt, etc - aber diese Merkmale reichen nicht aus, um realistische Simulationen der tatsächlichen kognitiven Aktivität zu ermöglichen. Die Maschinen benötigten Minuten bis Stunden, um zu berechnen, was die biologische "Maschine" im Bruchteil einer Sekunde schafft. Und das vor dem Hintergrund der nicht in Frage gestellten Annahme, dass wegen des Church-Turing-Theorems ein landläufiger Computer alles berechnen kann, was kognitiv möglich ist. Dabei sei es ein verstörender Befund, dass das Gehirn eine "clock frequency" von größenordnungsmäßig mehr als 100 Hz hat bei einer typischen Signalgeschwindigkeit von 10 m/sec - das ist nicht schneller als ein Fahrradfahrer. Also 10^2 Hz versus 10^{19} Hz und 10 m/sec versus 10^8 m/sec. Der Computer hat einen Geschwindigkeitsvorteil von $10^7 \times 10^7 = 10^{14}$ – also von 14 Größenordnungen, und doch ist das Hirn schneller als jeder Computer. Die Hirne aller Kreaturen auf der Erde funktionieren grundsätzlich anders als ein Computer mit der Standard-von Neumann-Architektur. Ein Teil des Unterschieds liegt in der massiven Parallelverarbeitung von Informationen im Gehirn; die bewirkt, dass Trillionen einzelner "Berechnungen" an den 10^{14} synaptischen Verbindungen gleichzeitig ablaufen, – und das kann zwischen 10 und

[75] der englische Ausdruck 'Folk Psychology' wird hier beibehalten, die deutsche Entsprechung wäre wohl 'Alltagspsychologie'.

100 mal pro Sekunde wiederholt werden. Die 100 Trillionen synaptischen Basisverbindungen sind die Grundelemente für die Berechnungen. Simultan wird ein ebenso paralleler Zugriff auf das Gedächtnis ermöglicht, und dieser Zugriff ist Teil des Gesamtprozesses, anders als im Computer, bei dem Informationsverarbeitung und Gedächtnissuche getrennte Prozesse sind. Funktionalisten schätzen den Wert dieser neurobiologischen Grundlagenforschung als gering ein, andererseits legen diese Ergebnisse doch sehr nahe, dass die Grundannahmen der Funktionalisten falsch sind. Churchland stellt die MR-These selbst nicht in Frage: sie sei ohne Zweifel richtig, – schon weil bestimmte Hirnteile die Aufgaben von anderen Hirnteilen zumindest partiell übernehmen können, wenn die anderen krankheits- bzw unfallbedingt ausfallen. Als er dieses Bekenntnis schrieb kannte er die neueren Arbeiten, die hier im Abschnitt 2.4.3.3 beschrieben wurden, noch nicht. Aber die Schlußfolgerung, dass die MR-These die Nichtreduzierbarkeit kognitiver auf neuronale Prozesse impliziere, sei "in fact wildly fallacious". Er gibt eine Reihe von allgemeinen, also nicht auf die Neurowissenschaft beschränkten Beispielen für multiple Realisierbarkeit:
Klang ist ein Phänomen auf molarem Niveau, weil Klang nur dann entstehen kann, wenn eine große Zahl von Partikeln in bestimmter Weise miteinander interagieren: in der Luft wie in allen Gasen, aber auch in Flüssigkeiten (Schiffsgeräusche, der Gesang von Walen), aber auch in "fester" Materie werden Schallwellen weitergeleitet. Aber überall findet die Weiterleitung als Kompressionswelle statt, wobei die Elemente des Mediums hin- und herschwingen. Die Geschwindigkeit, mit der sich die Wellen fortpflanzen hängt vom Medium ab. Gleichwohl ist die Akustik keine, wie Fodor sagen würde, 'autonome Wissenschaft', die Gesetze auf "ihrem eigenen Niveau von Beschreibung" zu finden sucht. Analoge Befunde ergeben sich, wenn die Thermodynamik betrachtet wird: sie ist ebenfalls keine "autonome Wissenschaft" mit ihrem "eigenen Niveau" von Beschreibungen. Die Reihe von Beispielen läßt sich fortsetzen. Und stets kann aus der multiplen Realisierbarkeit nicht gefolgert werden, dass Beschreibungen auf dem phänomenologischen Niveau nicht reduziert werden können, die Behauptung (5) sei "schlicht falsch" (Churchland, p. 41).

So direkt kann man es sagen. Es sei noch angefügt, dass keine zwei Menschen dieselbe Konfiguration von synaptischen Verbindungen haben. Jeder Mensch hat der Größenordnung nach 10^{14} synaptische Verbindungen, von denen die große Mehrheit erst nach der Geburt entsteht. Darüber hinaus kommt es bei jedem Menschen z.B. aufgrund von Lernprozessen (Neuroplastizität) zu individuellen Veränderungen der Stärke der Verbindungen. Die kognitiven Prozesse zweier verschiedener Menschen werden also im strengen Sinne nie identisch sein, – aber sie können ähnlich sein (s. p. 48/49). Für Vertreter der MR-These, die postulieren, dass die multiple Realisierbarkeit impliziert, dass mentale Ereignisse gar nichts mit der neuronalen Aktivität zu tun haben, haben diese neurophysiologischen Befunde keine Bedeutung. Aber diese Interpretation resultiert u.a. aus einer recht unorthodoxen Interpretation des

Begriffs der Disjunktion (7), Seite 22, und abgesehen davon ist es eine offene Frage, worin denn eigentlich der Erkenntnisgewinn einer dualistischen Position liegt.

ffff

2.4.5 Ausblick

Wenn, nach Putnam, alles eine TM ist (jede Berechnung kann durch eine TM durchgeführt werden), so ist auch der Geist (im Sinne von engl. mind) eine TM. Aber ein solcher Pankomputationalismus sollte nicht nur behauptet, sondern auch bewiesen werden, und Putnam hat ihn nicht bewiesen. Pankomputationalismus ist wahr nur in trivialen und philosophisch uninteressanten Fällen. Denn Verhalten kann i.A. nur approximativ durch eine TM beschrieben werden, d.h. als Approximation, und das Verhalten von komplexen Systemen *divergiert exponentiell* von jeder komputationalen Simulation (Strogatz (1994)). Strogatz argumentiert, dass die meisten Dinge keine zu berechnenden *Funktionen* haben (Beispiel: Hurricanes), und wenn doch, dann impliziert ihre Berechnung nicht die Berechnung relevanter *strings* in Übereinstimmung mit allgemeinen Regeln (etwa random number generators). Darüber hinaus sei die Frage, ob alle möglichen berechnenden Mechanismen Turing-vergleichbare Funktionen sind, noch offen (s. Piccinini 2003a), Copeland 2000, 2002). Fodors (1968a,b) Behauptung, psychologische Theorien könnten als Computerprogramme konzipiert werden, ist nie begründet(Piccinini (2004)), sondern allenfalls durch Plausibilitätsbetrachtungen suggeriert worden.

Aber das Gehirn ist — zumindest –kein digitaler Computer, wie u.a. Stephen Grossberg gleich im Vorwort seinem 2021 erschienenem *opus magnum Conscious Mind – Resonant Brain* feststellt, und allein die Übersicht über einige Eigenschaften des Gehirns am Ende des vorangegangenen Abschnitts 2.4.4.4 spricht für Grossbergs Dictum. Es ist nicht klar, warum man die Interpretation empirischer Daten, die sich aus Untersuchungen der Hirnaktivität ergeben, in das Prokustesbett einer Beschreibung im Rahmen von Turingmaschinen pressen soll, wie es die Rahmentheorie des komputationalen Funktionalismus ja implizit fordert, und umgekehrt ergeben sich aus dieser Theorie kaum Impulse für die Hirnforschung. Wenn man schon eine Rahmentheorie sucht, so könnte die Theorie der dynamischen Systeme in Frage kommen.

Dynamische Systeme: Allgemein ist ein System eine Struktur von miteinander wechselwirkenden Komponenten. Diese Wechselwirkungen sind Prozesse in der Zeit, die als kontinuierliche Größe angenommen wird. Dies bedeutet, dass Differentialgleichungen bzw. Systeme von Differentialgleichungen die natürliche Form der Beschreibung der Prozesse sind. Im Rahmen der Diskussionen über den funktionalen Komputationismus hat u.a Tim van Gelder (1998) die *Dynamische Hypothese* vorgeschlagen. Damit meinte er, neuronale Teilpopu-

lationen oder gleich das ganze Gehirn als dynamisches System zu konzipieren, womit auch die Brücke zu Beschreibungen mentaler Prozesse durch Differentialgleichungen geschlagen wird. Der Begriff des dynamischen Systems ist allgemeiner als die Theorie der Differentialgleichungen, wenn auch die Titel von Lehrbüchern der Mathematik die begriffliche Äquivalenz nahelegen[76]. Auf naturwisssenschaftlicher Seite bedurfte es derartiger Betrachtungen gar nicht, schon weil insbesondere Überlegungen über eine möglicherweise dualistische Natur der Kognitionen als eher abwegig angesehen werden und die theoretische Analyse von neuronalen Systemen nachgerade automatisch auf die Anwendung von Differentialgleichungen führt: (Dayan & Abbot (2001), Gerstner & Kistler (2002). Das liegt einfach daran, dass Wechselwirkungen zwischen Variablen in natürlicher Weise durch Differentialgleichungen beschrieben werden, wenn nur gewisse, recht allgemeine Voraussetzungen erfüllt werden (Grebogi, C., Ott, E., & Yorke, J. A. (1987)). Die zentrale Voraussetzung ist die Stetigkeit und Differenzierbarkeit der Variablen, die betrachtet werden. Natürlich kann man fragen, wozu überhaupt eine mathematische Modellierung der Prozesse und "Mechanismen" gut sein soll. Die allgemeine Antwort darauf ist, dass bereits gewöhnliche Datenanalysen, z. B. die von EEG-Daten[77] auf der Annahme bestimmmter Modelle beruht, wie überhaupt alle deskriptiv- und inferenzstatistischen Analysen auf der Annahme von Modellen beruhen. Spezielle Modelle für bestimmte kognitive Prozesse oder die Interaktion von Emotionen haben den den Zweck, sich genauere Vorstellungen über diese Prozesse machen zu können, die sozusagen umgangssprachliche Beschreibung von Prozessen kann spezifische Aspekte nicht erfassen, – eben dieser Sachverhalt hat ja zur Entwicklung der formalen Sprache der Mathematik geführt.

Der Begriff der Berechnung ist nicht auf die turingsche Berechenbarkeit beschränkt. Schon früh und ohne große philosophische Argumentationen ist bei biologischen Fragestellungen vom Begriff des dynamischen Systems Gebrauch gemacht worden, etwa beim Lotka-Volterra Modell der Populationsdynamik (1925, 1926). Weiter hat Nicolas Rashevsky (1899 – 1972) biologische Prozesse als Aktivität von dynamischen Systemen beschrieben, ohne diesen Ansatz jedoch in Bezug auf experimentell gewonnene Daten zu diskutieren. In engem Zusammenhang mit der Empirie hat der niederländische Ingenieur H. DeLange (1958) das visuelle System als ein dynamisches System betrachtet, dessen Struktur sich durch Messungen der Wahrnehmbarkeit von sinusförmig modulierten Stimuli (im Zeit- und im Ortsbereich) entschlüsseln lässt, – zumindest im Prinzip. Wie sich zeigen läßt erlaubt es die Kenntnis der Reaktionen auf sinusförmig modulierte Stimuli, die Reaktion auf beliebige Stimuli vorherzusagen, und darüber hinaus erlauben derartige Messungen die Analyse der funktionalen Struktur der jeweils involvierten neuronalen Mechanismen. DeLanges

[76] z.B. Arrowsmith, D.K, Place, C.M.: An Introduction to Dynamical Systems. Cambridge University Press 1990; das Buch fokussiert auf die Theorie der Differentialgleichungen.

[77] z.B. multivariate Zeitreihenanalyse, Karhunen-Loève-Entwicklung

Ansatz wurde im Bereich der Psychophysik Standard. Von 'Berechnungen' des System war allenfalls umgangssprachlich die Rede. Erst David Marr (1945 - 1980) hat den Begriff 'Computation' in seinem Buch *Vision* (1982) sozusagen offiziell gemacht. Marr ging ebenfalls von einer systemtheoretischen, durch neuroanatomische und -physiologische Befunde motivierte Repräsentation des visuellen Systems aus und beschrieb Aspekte der visuellen Wahrnehmung durch Eigenschaften der durch Stimulusmuster erzeugten Aktivitätsverteilung. Denkt man in Termen des zugrundegelegten Modells, so kann man sagen, dass das visuelle System seine Reaktionen auf Stimulusmuster "berechnet", wobei dieser Ausdruck eher im übertragenen Sinn gebraucht wird, – ein Modell *repräsentiert* die Aktivitäten von neuronalen Populationen.

Konnektionismus: Natürlich kann der dynamische Ansatz nicht nur auf neuronale, sondern allgemein auf "höhere" psychologische Dynamiken angewendet werden. Generell besteht eine enge Beziehung des Dynamik-Ansatzes zum Konnektionismus (Rumelhart & Hinton (1986)[78], und Rumelhart, Hinton, & Williams (1986), und schließlich Neuronaler Netzwerke (Tanaka et al.(2021)) . Hava Siegelmann (1995) konnte zeigen, dass neuronale Netze konzipiert werden können, die nicht mehr den Begrenzungen des zu Gödels (1931) Theoremen korrespondierenden Turingschen Halte-Problems unterliegen.

Um den Begriff des dynamischen Systems zu illustrieren kann die Dynamik von Emotionen und Affekten betrachtet werden, zumal Emotionen mit rein kognitiven Aktivitäten interagieren, – kognitive Ativitäten helfen u.a. bei der Kontrolle von Affekten, und emotionale Zustände können Einfluß auf den Ablauf von Kognitionen haben. In der langen Geschichte der Emotionsforschung ist schon früh die Hypothese formuliert worden, dass Emotionen Kombinationen von bestimmten Basisemotionen sind. Dabei haben sich zwei alternative Ansätze herausgeschält (Lindquist et al. (2012)):
(1) Die *Lokations*-Hypothese[79], bei der angenommen wird, dass bestimmte Emotionen zu bestimmten Hirnregionen korrespondieren. So wird etwa die Amygdala aktiviert, wenn Furcht, Ärger oder Ekel erlebt werden. Die Lokations-Hypothese erinnert an die Identitätstheorien von Place und Smart aus den 50er Jahren.
(2) Die *psychologisch-konstruktivistische* Hypothese, derzufolge sich Emotionen aus basaleren psychologischen Operationen ergeben, die in fast allen mentalen Aktivitäten eine Rolle spielen. Auch dieser Ansatz ist nicht grundsätzlich neu, er findet sich schon bei William James (1842 – 1910) und Wilhelm Wundt (1832 – 1920).

Lindquist et al. unterzogen eine Vielzahl von fMRI-Untersuchungen einer Metaanalyse. Der grundsätzliche Befund ist, dass die Daten eher für die Hypo-

[78] Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. *Parallel distributed processing: Explorations in the microstructure of cognition,* 1(45-76), 26.
[79] locationist approach

these (2) sprechen: es gibt stets eine Menge verschiedener Hirnbereiche, die bei emotionalen und nicht-emotionalen, eher kognitiven Erfahrungen miteinander interagieren. "'Diskrete", d.h. deutlich unterschiedliche Emotionen wie Furcht, Ärger etc ergeben sich demnach eher als Kombinationen basalerer Prozesse. Wenn also im Folgenden von Frustration und Aggression die Rede ist, so soll damit nicht unterstellt werden, dass es sich dabei um eigenständige Kategorien handelt; die Rede ist dann auch von der Einbettung dieser Emotionen bzw. Affekte in ein übergeordnetes Netzwerk.

Um ein dynamisches System explizit vorzustellen, werde die von Dollard et al. (1939) vorgestellte These, dass der Entwicklung von Aggression stets eine Frustration vorausgehe diskutiert. Das hier vorgestellte dynamische Modell ist ein stark vereinfachtes "toy example", also ein Spielzeugbeispiel, das nicht die gesamte Komplexität der Interaktion zwischen Emotionen darstellen soll, sondern nur einige grundsätzliche Eigenschaften dynamischer Systeme und der sich daraus ergebenden Konsequenzen für die empirische Forschung illustrieren soll.

Es wird angenommen, dass die Intensität der Emotionen Frustration und Aggression durch Funktionen $F = F(t)$ und $A = A(t)$ der Zeit t repräsentiert werden kann, wobei diese Funktionen gewissermaßen glatt sind in dem Sinne, dass sie weder Sprünge noch Knicke haben, weil bei Sprüngen oder Knicken keine Differentialquotienten definiert sind. Es existiere für alle t im Beobachtungsintervall $I_T = [0, T]$ der Differentialquotient

$$\frac{dF(t)}{dt} = \dot{F}(t) = \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t},$$

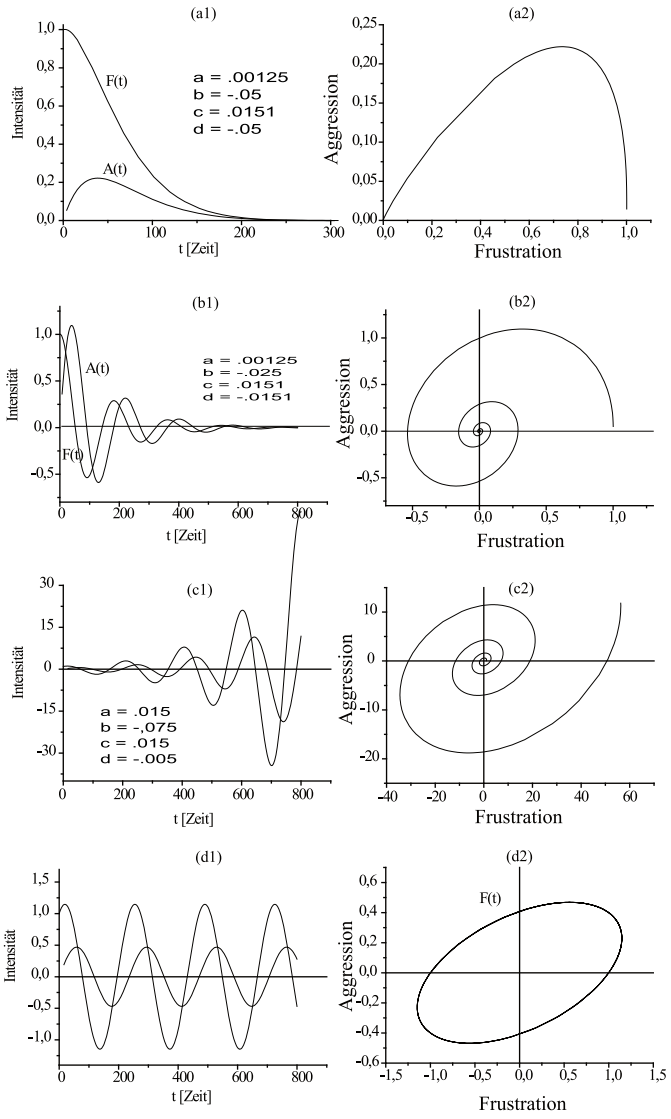
eine analoge Aussage gelte für $A(t)$; $\dot{F}(t)$ ist eine Kurzsschreibweise für dF/dt. Die Differentialquotienten $\dot{F}$ und $\dot{A}$ sind Maße für die Veränderung von F und A zur Zeit t. Die Beziehung zwischen einer Funktion und dem Differentialquotienten ist eindeutig bis auf eine additive Konstante. Bei Differentialgleichungen wird dieser Sachverhalt ausgenutzt, um aus Annahmen über die Veränderung einer oder mehrer Funktionen Ausdrücke über diese Funktionen zu erhalten. Die beiden folgenden DGLn sind eines der einfachsten Systeme von gekoppelten Differentialgleichungen:

$$\frac{dF(t)}{dt} = \dot{F} = a_{11}F(t) + a_{12}A(t) + s_1(t) \tag{18}$$

$$\frac{dA(t)}{dt} = \dot{A} = a_{21}F(t) + a_{22}A(t) + s_2(t) \tag{19}$$

Das Paar $[F(t), A(t)]$ definiert den *Zustand* des Systems zum Zeitpunkt t. Offenbar ist ein Zustand stets eine Mischung der beiden Emotionen oder Affekte, die sich mit der Veränderung von t verändert. Dabei kann für einen gegebenen Verlauf der Frustration die Aggression einen nur schwachen oder auch ausgeprägteren Verlauf haben; im ersteren Fall ist auch von Mikroaggression die Rede.

Abbildung 1: Typen von Reaktionen; (d1) und (d2): $a_{11} = a = .015$, $a_{12} = b = -.075$, $a_2 1 = c = .0125$, $a_{22} = d = -.015$. Die Typen sind durch bestimmte Beziehungen zwischen den Parametern bestimmt. Innerhalb jeden Typs existieren beliebig viele Varianten des Typs.



Tragt man für jeden Wert von t den Wert von A gegen den von F auf, so ensteht eine Kurve, die den Verlauf der Zustände abbildet; die Rede ist auch von einer "Trajektorie" des Systems. Eine Trajektorie hängt von den Koeffizienten a_{ij} ab. s_1 und s_2 sind "Störungen" des Systems, die natürlich ebenfalls die Trajektorien des Systgems mitbestimmen. Die Gleichungen (18) und (19) repräsentieren ein lineares System, weil F und A nur additiv in die Gleichun-

gen eingehen[80]. Im Allgemeinen sind biologische Systeme nichtlinear, auf der rechten Seite der Gleichungen stehen dann zum Beispiel Produkte $F(t)A(t)$. Derartige Terme haben die Bedeutung, dass etwa $A(t)$ sich direkt proportional zu $F(t)$ verändert. Lineare Systeme sind dann Approximationen an das System für kleine Auslenkungen des Systems; die Diskussion der Dynamik für kleine Auslenkungen ist wichtig, um das Verhalten des Systems in der Nachbarschaft von Gleichgewichtspunkten zu verstehen, die entweder stabil oder instabil sein können und die deshalb wichtig für das Verständnis des Gesamtsystems sind.

Die Koeffizienten a_{12} und a_{21} sind die *Kopplungskoeffizienten*, die die Wechselwirkung zwischen F und A bestimmen. Für $a_{12} = 0$ wirkt die Aggression nicht auf die Frustration zurück, und für $a_{21} = 0$ hat die Frustration keinen Einfluß auf die Aggression. Die a_{ij} können vielleicht für kurze Zeitintervalle in guter Näherung konstant sein, aber sie werden im Allgemeinen ebenfalls Funktionen der Zeit sein; sie repräsentieren gewissermaßen die Einbettung dieses Teilsystems in die allgemeine emotionale Dynamik, womit die beiden Gleichungen eine sehr allgemeine Dynamik repäsentieren, obwohl sie linear sind, d.h. die rechten Seiten die Veränderungen $\dot{F}$ und $\dot{A}$ nur additiv beeinflussen. Für $s_1 = s_2 = 0$ ist das System *autonom*. Das Verhalten des autonomen System kann als Reaktion auf eine kleine "Auslenkung" des System gesehen werden, wenn das System nach der Auslenkung keinen weiteren Input mehr hat, es ist die Reaktion auf einen Impuls, der selbst keine zeitliche Ausdehnung hat[81]. Die Impulsantworten eines System sind charakteristisch für das System, der Aufbau des Systems kann im Prinzip aus ihnen abgeleitet werden.

Für $\dot{F} = 0, \dot{A} = 0$ finden keine Veränderungen von F und A statt, das System ist in einer Gleichgewichtslage; die Rede ist auch von einem Fixpunkt des Systems. Offenbar definiert $F = A = 0$ eine solche Lage, allerdings nicht notwendig die einzige: abhängig von den Werten der Koeffizienten a_{ij} kann es noch Paare $(F_k, A_k))$ geben mit $F_k \neq 0$, $A_k \neq 0$, für die die rechten Seiten des Systems ebenfalls gleich Null sind. Für den Fall der Stabilität einer Gleichgewichtslage kehrt das System nach kleinen Störungen in diese Lage zurück, der Fixpunkt ist ein "Attraktor". Für instabile Gleichgewichtslagen kann es sich von der Gleichgewichtslage entfernen und in den Attraktorbereich eines anderen Fixpunkts geraten, oder es schaukelt sich unbegrenzt auf, indem es immer größer werdende Werte für F und A generiert. Da das System in ein Gesamtsystem eingebettet ist kann man davon ausgehen, dass sich die a_{ij} verändern und nur während einer Auslenkung annähernd konstant sind; man betrachtet

[80] Diese Definition von 'linear' ist eher anschaulich gemeint, die formale Definition für ein lineares System L ist: L ist *linear*, wenn für alle t $L(ax + by) = aL(x) + bL(y)$ gilt, wenn x und y die betrachteten Funktionen sind.

[81] Diese Idee scheint unrealistisch zu sein, und tatsächlich ist sie eine mathematischen Idealisierung, die auf den britischen Pphysiker Paul Dirac (1902 – 1984) zurückgeht: ein Dirac-Impuls ist (i) eine Funktion $\delta(t - t_0) = 0$ für alle $t \neq t_0$, und $\delta(t - t_0) = \infty$ für $t = t_0$, und (ii) die Fläche unter $\delta(t - t_0)$ ist gleich 1. Dirac δ-Funktionen ergeben sich asymptotisch z.B. aus Fluktuationen wie der Gauss-Dichte $f(t) = (1/\sigma\sqrt{2\pi}) \exp(-t^2/(2\sigma^2)$ für $\sigma \to 0$.

dann das Verhalten des Systems für die Werte der Koeffizienten, die es bei der Auslenkung gerade hat. Man erhält auf diese Weise eine approximative Darstellung des Verhaltens.

Das Modell erklärt, warum die Dynamik des Systems ((18) , (19)) praktisch nicht vorhersagbar ist, obwohl die Gleichungen deterministisch sind, wenn man nichts über die die Abhängigkeit der Werte der a_{ij} von der Umgebung des Systems weiß. So ist es möglich, dass sich das Verhalten des Systems in unvorhergesehener Weise qualitativ verändert, wie noch illustriert werden wird.

Typen von Reaktionen: Tatsächlich kann man von qualitativ verschiedenen Typen von Reaktionen sprechen. Dazu werde angenommen, dass die a_{ij} als Funktionen der Zeit langsam variieren relativ zur Geschwindigkeit der Veränderung der Antwort des Systems auf eine Störung. So gelte $a_{ij} \approx$ konstant für die Dauer der Reaktion auf einen "Input". Für lineare Systeme mit konstanten Koeffizienten genügt es, die Reaktion auf einen Impuls zu kennen, wobei ein Impuls eine Störung von sehr kurzer Dauer ist. Die Antwort des Systems heißt deswegen *Impulsantwort*. Es kann gezeigt werden, dass (i) die Impulsantworten charakteristisch für das System sind, und (ii) die Reaktion auf einen Input $s(t)$ von beliebigem zeitlichen Verlauf unter Verwendung der Impulsantwort berechnet werden kann. Auf die mathematischen Details soll hier nicht eingegangen werden kann, das Wesentliche ist, dass die Impulsantworten das System für die Dauer der approximativen Konstanz der a_{ij} charakterisieren.

Abbildung 1 zeigt die Reaktionen des Systems für verschiedene Kombinationen von a_{ij}-Werten. Die linke Spalte zeigt die Verläufe von $F(t)$ und $A(t)$ als Funktionen der Zeit, die rechte Spalte zeigt die zu diesen Verläufen korrespondierenden *Phasendiagramme*: zu jedem Zeitpunkt t haben ja F und A bestimmte Werte, und man kann für jeden Wert $F(t)$ den zugehörigen $A(t)$-Wert abtragen. Die so entstehende Kurve ist das Phasendiagramm. In (a1) nimmt $F(t)$ für $t = 0$ einen maximalen Wert ($F = 1$) an, und es ist $A = 0$. F strebt dann gegen Null (etwa bei $t = 200$), während A zuerst rasch ansteigt, um denn ebenfalls gegen Null bei ca $t = 200$ zu streben. Man muß die Abbildung (a2) gewissermaßen von rechts nach links lesen. (b1) unf (b2) zeigen einen oszillatorischen Verlauf von F und A; A ist ein wenig nachläufig, und (b2) zeigt, wie die Reaktionen $F(t)$ und $A(t)$ mit wachsendem t gegen Null "spiralen". Das System ist wie in (a1, a2) stabil, nach einer Auslenkung kehrt es in die Gleichgewichtslage $(F, A) = (0, 0)$ zurück. (c1) un (c2) zeigen den Fall eines im Vergleich zu (b1) und (b2) umgekehrten Verlaufs: F und A haben für kleine Werte von t nur kleine Werte, die sich dann aber mit wachsendem t aufschaukeln; das System ist instabil. In (d1 d2) wird ein weiterer möglicher Verlauf gezeigt: Die Störung durch einen Impuls versetzt das System in eine dauerhafte Oszillation von F- und A-Werten. Die vier Muster von Reaktionen auf eine Störung repräsentieren Typen von Reaktionen: kleine Abweichungen von den angegebenen Werten der a_{ij} erzeugen ähnliche Verläufe von F und

A. Am plausibelsten sind wohl Verläufe wie in (a1, a2): kleine Frustrationen erzeugen Mikroaggressionen (vergl.inbesondere Abbildung 2), und die Intensitäten beider Emotionen gehen schnell gegen Null. Der Typ (c1, c2) (Abbildung 1) zeigt Verläufe, die mit Kontrollverlust einhergehen können, – schon eine kleine Auslenkung kann eine Tobsuchtsanfall hervorrufen. Grundsätzlich ist es so, dass der Gesamtzustand des Systems zu einem gegeben Zeitpunkt t durch die Werte von F und A zu diesem Zeitpunkt gegeben ist, eine Person wird zu jedem Zeitpunkt eine Mischung der beiden Emotionen (oder Affekte) erleben, die sich ständig verändert.

Die Abbildung 1 zeigt die wesentlichen Klassen von Verhaltensweisen eines Systems der Art ((18), (19)); eine vollständige Analyse derartiger Systeme findet man in Arrowsmith & Place (1982)[82], p. 52. Aber auch innerhalb eines Typs gibt es beliebig viele Variationen; Abb. 2 zeigt zwei unterschiedliche Fälle für den Fall a1 in Abb. 1. Der Fall a1 zeigt für einen gegeben Verlauf der Frustration eine nur kleine Aggressionsreaktion – man könnte von einer Mikroaggression sprechen. a2 zeigt ds Phasendiagramm. Der Fall b1 zeigt einen Verlauf der Frustration, der dem im Fall a1 sehr ähnlich ist, bei dem der Verlauf der Intensität der Aggression sehr ausgeprägt ist b2 zeigt das zugehörige Phasendiagramm. Die Unterschiede zwischen den beiden Fällen ergeben sich aus den unterschiedlichen Gesamtzuständen, die sich in unterschiedlchen Parameterwerten manifestieren.

Grundsätzlich zeigen die verschiedenen Typen interne Abläufe, wie sich diese auf das Verhalten, d.h. auf die Handlungen einer Person auswirken ist eine andere Frage. Das hier vorgestellte Modell enthält keinerlei Annahmen etwa über Kontrollfunktionen des Frontalhirns, ob eine Person, die einen starken Ausschlag der Aggressionsvariable erlebt diesen auch handlungsrelevant werden läßt ist keineswegs gesagt. Dies stellt für eine experimentelle Überprüfung der F-A-Hypthese eine große Schwierigkeit dar (vergl. dazu Breuer & Elson (2017)). Soll eine Versuchsperson die erlebte Intensität einer Emotion zum Beispiel auf einer Rating-Skala anzeigen, so ist keineswegs klar, wie sie diese Abbildung aus den erlebten Verläufen gewinnt. Dass derartige Daten dann nicht viel über die internen Prozessse aussagen ist nicht verwunderlich. Dabei gilt die allgemeine Einsicht:

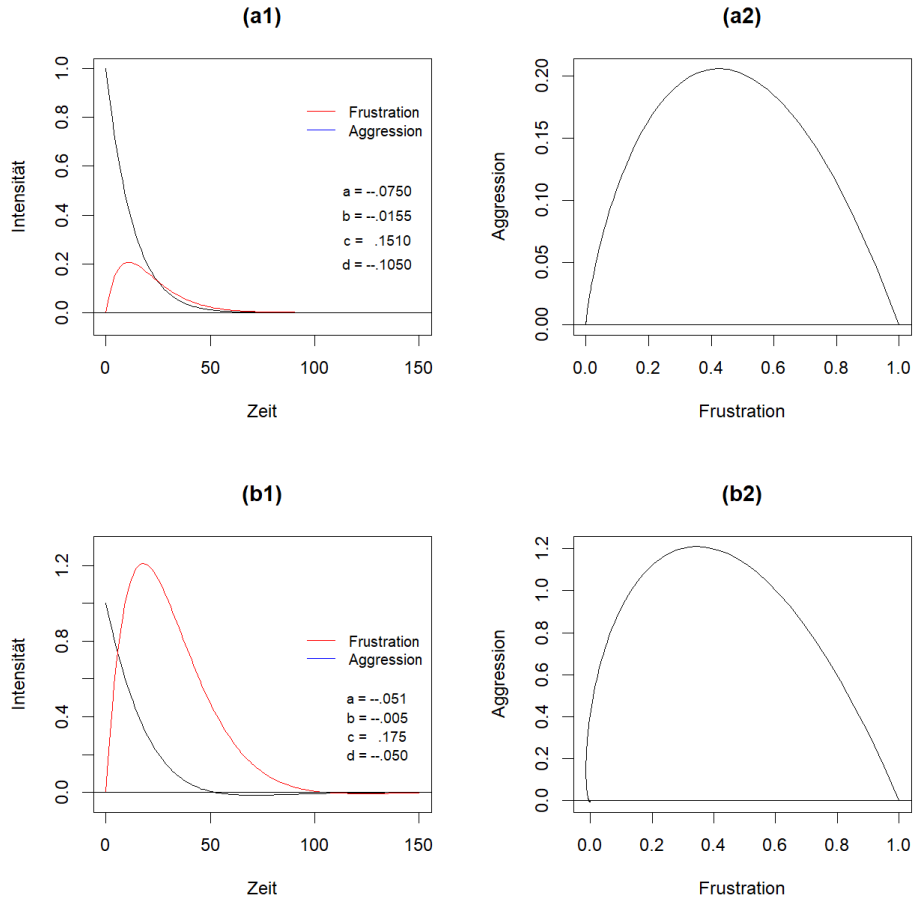
Alle Modelle sind falsch, aber einige sind nützlich.

Diese Einsicht wird dem britischen Statistiker George Edward Pelham Box ((1976) zugeschrieben, der an William of Occams Rat erinnerte, dass eine möglichst ökonomische Beschreibung der Phänomene gesucht werden sollte[83]. Modelle wie [(18), (19)] verdeutlichen den Sachverhalt, dass bei Kognitionen

[82] Arrowsmith, D.K., Place, C.M.: Ordinary Differential Equations, Chapman & Hall, London 1982

[83] "Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical

Abbildung 2: Typen von Reaktionen;$a = a_{11}$, $b = a_{12}$, $c = a_{21}$, $d = a_{22}$



sehr wohl enge Zusammenhänge zwischen verschiedenen Variablen existieren können, die in der normalen, d.h. in der Umgangssprache nicht präzise beschrieben werden können, – deshalb ist die formale Sprache der Mathematik ja entwickelt worden. Die Fülle der allein von Lindquist et al. zusammengetragenen empirischen Befunde legt jedenfalls nahe, dass die Produktionsstätte mentaler Ereignisse das Gehirn ist (Zur Beziehung zwischen Dynamischen Systemen und Neuronalen Netzen vergl. Kumpati & Kannan (1990).

description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity." Box (1976), und "For such a model there is no need to ask the question "Is the model true?". If "truth" is to be the "whole truth" the answer must be "No". The only question of interest is "Is the model illuminating and useful?" (Box(1979), see Box, G. E. (1979). Robustness in the strategy of scientific model building. In *Robustness in statistics*, pp. 201-236), Academic Press)

3 Anhang

Überabzählbarkeit der reellen Zahlen: Es sei $\{z_i\}$ eine beliebige Folge von reellen Zahlen aus $(0,1) = \{x|0 < x < 1\}$. Es existiert mindestens eine reelle Zahl, die nicht in $\{z_i\}$ vorkommt.

Eine beliebige Zahl aus dem Intervall $(0,1)$ ist durch eine Folge von Dezimalzahlen bestimmt, also etwa $.347191\ldots$; die Anzahl der Nachkommazahlen kann endlich oder unendlich sein, im Falle unendlich vieler Nachkommmazahlen kann die Folge periodisch oder aperiodisch sein. Die Folge $\{z_i\}$ kannn dann in der Form

$$\begin{aligned} z_1 &= .v_{11}v_{12}v_{13}\ldots \\ z_2 &= .v_{21}v_{22}v_{23}\ldots \\ z_3 &= .v_{31}v_{32}v_{33}\ldots \\ &\vdots \end{aligned}$$

angeschrieben werden. Georg Cantor[84] konnte mit dem von ihm entwickelten *Diagonalverfahren* zeigen, dass die Anzahl der reellen Zahlen größer ist als die der natürlichen Zahlen, d.h. die Menge der rellen Zahlen ist *überabzählbar*. In der Tabelle 1 erscheinen die z_i in den Spalten: so ist z_1 durch die Dezimalzahlen v_{11}, v_{21} etc definiert: der erste Index bezeichnet die Position der Dezimalzahl in der Folge für z_1, der zweite Index kennzeichnet die Dezimalzahl, etwa z_1. Es sei

Tabelle 1: Cantors Diagonalverfahren; v_{ij} i-te Dezimalstelle, j-te Zahl

	z_1	z_2	z_3	$\cdots$
1	v_{11}	v_{12}	v_{13}	$\cdots$
2	v_{21}	v_{22}	v_{23}	$\cdots$
3	v_{31}	v_{32}	v_{33}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$

nun $x = .x_1x_2x_3\ldots$ eine reelle Zahl aus $(0,1)$. Die x_i werden wir folgt definiert: Zu jeder Dezimalzahl x_i von x kann die Zahl z_i der Folge $\{z_i\}$ betrachtet werden: Wenn $v_{ii} = 5$, so wird $v_{ii} = 4$ gesetzt, sonst $v_{ii} = 5$. Dann folgt $x \neq z_i$. So geht man durch die ganze Folge $\{z_i\}$ und erhält eine Zahl x, die sich von allen Zahlen der Folge $\{z_i\}$ in mindestens einer Diagonalstelle unterscheidet und für die $0 < x < 1$ gilt. x die *Dezimalzahl*, die der Folge $\{z_i\}$ zugeordnet wird.

Damit enthält $\{z_i\}$ nicht *alle* reellen Zahlen aus $(0,1)$. *Für jede Folge* $\{z_i\}$ *von Zahlen aus* $(0,1)$ *existiert eine Zahl aus* $(0,1)$*, die nicht in der Folge*

[84]Georg Ferdinand Ludwig Philipp Cantor (1845 – 1918) war ein deutscher Mathematiker, der insbesondere als Begründer der Mengenlehre bekannt geworden ist. Mit seiner Theorie der transfiniten Ordnungszahlen hat er "Ordnung im Unendlichen" geschaffen.

enthalten ist. Da die ausgewähte Folge beliebig gewählt werden kann folgt. dass keine Folge alle Zahlen aus $(0, 1)$ enthält.

Fasst man Folgen als Abbildungen $\mathbb{N} \to (0, 1)$ auf, so bedeutet das Ergebnis eben, dass es keine umkehrbar eindeutige Abbildung $\mathbb{N} \to (0, 1)$ gibt. Deshalb ist $(0, 1)$ nicht gleichmächtig zu $\mathbb{N}$, d.h. $(0, 1)$ ist überabzählbar.

Literatur

[1] Arrowsmith, D.K, Place, C.M.: An Introduction to Dynamical Systems. Cambridge University Press 1990

[2] Atkinson, R.C., Shiffrin, R.M.(1968) Human Memory: a proposed system and its control processes. In: K.W. Spence andJ.T. Spence: the Psychology of Learning and Motivation, Vol. 2. New York 1968

[3] Bechtel, W., McCauley, R. (1999), Heuristic Identity Theory (or Back to the Future): the Mind-Body Problem Against the Background of Research Strategies in Cognitive Neuroscience, *Proceedings of the 21st Annual Meeting of the Cognitive Science Society*, Mahwah, NJ: Lawrence Erlbaum Associates

[4] Bechtel, W., Mundale, J. (1999) Multiple Realizability Revisited: Linking Cognitive and Neural States. *Philosophy of Science,* 66 (2), 175-207

[5] Beckermann, A.: Analytische Einführung in die Philosphie des Geistes. Berlin 2008

[6] Benacerraf, P. (1967). God, the devil, and Gödel. *The monist,* 9-32

[7] Berry, M. (1994) Asymptotics, Singularities and the Reduction of Theories. *Logic, Methodology and Philosophy of Science IX*, 597 – 607

[8] Bickle, J. (1992a) Mental Anomaly and the new Mind-Brain Reductionism. *Philosophy of Science,* 59(2), 217 – 230

[9] Bickle, J. (1992) Multiple Realizability and Psychophysical Reduction, *Behavior and Philosophy*, 20(1), 47-58

[10] Bickle, J. (1996) New Wave Psyhophysical Reductionism and the Methodological Caveats. *Phenomenological Research*, LVI (1), 57 – 76

[11] Bickle, J. (2003). Philosophy and neuroscience: A ruthlessly reductive account (Vol. 2). Springer Science & Business Media.

[12] Bickle,J. (2006) Reducing mind to molecular pathways: explicating the reductionism implicit in current cellulbasalenar and molecular neuroscience. *Synthese*, 151, 411-434

[13] Bickle, J.: Multiple Realizability, *Stanford Encyclopedia of Philosophy*, 2020

[14] Block, N.: Functionalism. *The Encyclopedia of Philosophy Supplement.*

[15] Bober-Irizar, M., & Banerjee, S. (2024). Neural networks for abstraction and reasoning: Towards broad generalization in machines. arXiv preprint arXiv:2402.03507

[16] Box, G. E. P. (1976), "Science and Statistics", *Journal of the American Statistical Association,* 71: 791–799

[17] Breuer, J., & Elson, M. (2017). Frustration-Aggression Theory. In P. Sturmey (Ed.), The Wiley Handbook of Violence and Aggression (pp. 1-12). Chichester: Wiley Blackwell

[18] Bridgman, P. W.: The Logic of Modern Physics. MacMillian, New York 1927

[19] Buechner, J. (2010). Are the Gödel incompleteness theorems limitative results for the neurosciences?. *Journal of biological physics,* 36(1), 23-44.

[20] Carnap,R., Neurath, O.: Wissenschaftliche Weltauffassung - der Wiener Kreis, in:

[21] Chalmers, D. J., The Conscious Mind: In Search of a Fundamental Theory, New York and Oxford: Oxford University Press 1996

[22] Churchland, P.M. (1985) Reduction, Qualia, and the direct Introspection of Brain States. *Journal of Philosophy*, 82(1)

[23] Churchland, Paul M. (2005) Functionalism at forty: A critical retrospective. *The Journal of Philosophy*, 102(1), 33-50

[24] Clapp, L. (2001). Disjunctive properties: Multiple realizations. *The Journal of Philosophy,* 98(3), 111-136.

[25] Clark, A., Chalmers, D. (1998) The Extended Mind, *Analysis,* 58, 10-23

[26] Chomsky, N. (1959) A Review of B. F. Skinner's Verbal Behavior. *Language*, 35, No. 1

[27] Copeland, B. J., Fan, Z. (2022). Did Turing Stand on Gödel's Shoulders?. *The Mathematical Intelligencer,* 44(4), 308-319.

[28] Cordeschi, R., & Frixione, M. (2007) Computationalism under attack. In: *Cartographies of the Mind: Philosophy and Psychology in Intersection*, (pp. 37-49) Dordrecht: Springer Netherlands

[29] Couch, M. B. (2004). Discussion: A Defense of Bechtel and Mundale. *Philosophy of Science*, 71(2), 198-204.

[30] Davidson, D. (1970) Mental Events, in: Actions and Events, Oxford: Clarendon Press, 1980.

[31] Dayan, P., Abbot, L.F.: Theoretical Neuroscience – Computational and Mathematical Modeling of Neuronal Systems. Cambridge 2001

[32] DeLange, Dzn H. (1958) Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light. *J Opt Soc Am.*, 48(11), 777–784 (Dzn steht für Dirks zoon, also Dirks Sohn)

[33] Denker, J.S., leCun, Y.: Natural versus "Universal" Probability, Complexity, and Entropy, *IEEE, Xplore* 2022

[34] Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., & Sears, R. R. (1939). Frustration and aggression. Yale University Press. https://doi.org/10.1037/10022-000

[35] Edelman, G.M., Tononi, G.: Gehirn und Geist – wie aus Materie Bewusstsein entsteht. München 2004

[36] Eliasmith, C. (2002), The myth of the Turing machine: The failings of functionalism and related theses. *Journal of Experimental & Theoretical Artificial Intelligence,* 14(1), 1-8

[37] Endicott, R. P. (1993). Species-specific properties and more narrow reductive strategies. *Erkenntnis*, 38(3), 303-321

[38] Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex*, 1(1), 1-47.

[39] Feyerabend, P.K. (1962) Explanation, Reduction, and Empiricism, *Minnesota Studies in the Philosophy of Science, III*, in: Feigl, H. Maxwell, G. (eds), Minneapolis 1962, 28-97

[40] Fischer-Baum, S., Jang, A., Kajander, D. (2017). The cognitive neuroplasticity of reading recovery following chronic stroke: A representational similarity analysis approach. *Neural plasticity,* 2017(1), 2761913.

[41] Fodor, J. A.: Psychological explanation: An introduction to the philosophy of psychology. New York: Random House 1968

[42] Fodor, J. A. (1968). The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy,* 65, 627–640.

[43] Fodor, J.A. (1974) Special Sciences (Or: The Disunity of Science as a Working Hypothesis), *Synthese*, 28(2), 97–115

[44] Francescotti, R. M. (1997). What multiple realizability does not show. *The Journal of Mind and Behavior,* 13-27.

[45] Frey, G.: Theorie des Bewußtseins. München 1980

[46] Gerstner, W., Kistler, W.: Spiking Neuron Models – Single Neurons, Populations, plasticity. Cambridge 2002

[47] Gödel, K. (1972). A philosophical error in Turing's work. *Kurt Gödel: Collected Works,* 2, 306.

[48] Godbey, J. W. (1978), Disjunctive predicates and the reduction of psychology. *Mind,* 87(347), 433-435.

[49] Graves, A. (2014). Neural Turing Machines. *arXiv preprint arXiv:* 1410.5401.

[50] Graves, A., Wayne, G., Danihelka, I. (2014). Neural turing machines. *arXiv preprint arXiv:*1410.5401

[51] Grebogi, C., Ott, E., Yorke, J. A. (1987). Chaos, strange attractors, and fractal basin boundaries in nonlinear dynamics. *Science,* 238(4827), 632-638.

[52] Grossberg, S.: Conscious Mind – Resonant Brain – How each Brain makes a mind. Oxford University Press, 2021

[53] Hebb, D.: The organization of behavior. A neuropsychological theory. New York 1949 (Neuauflage: Erlbaum Books, Mahwah, N.J. 2002)

[54] Hemmo, M., Shenker, O. (2015). The emergence of macroscopic regularity. *Mind & Society,* 14, 221-244.

[55] Hilbert, D. Ackermann, W. Grundzüge der Theoretischen Logik. Springer, Berlin, 1928.

[56] Hoffmann, D.W.: Grenzen der Mathematik. Spektrum Akademischer Verlag Heidelberg 2011

[57] Hooker, C. A. (1981). Towards a general theory of reduction. Part I: Historical and scientific setting. Part II: Identity in reduction. Part EQ: Cross-categorical reduction. *Dialogue*, 20, 38-59, 201-236, 496-529.

[58] Kemeny, J.G: Oppenheim, P (1956) On Reduction, *Philosophical Studies*, VII, 6 – 17

[59] Kim, J. (1972) Phenomenal properties, Psychophysical Laws, and identity Theory. *The Monist*, 56, 177 – 192

[60] Kim, S. (2002) Testing Multiple Realizability: A Discussion of Bechtel and Mundale, *Philosophy of Science*, 69 (4), 606-610

[61] Koons, R. C., and Bealer, G,. (eds): The waning of materialism. OUP Oxford, 2010.

[62] Kuhn, T.S., The structure of scientific revolutions. Chicago 1962

[63] Kumpati, S. N., Kannan, P. (1990) Identification and control of dynamical systems using neural networks. *IEEE Transactions on neural networks,* 1(1), 4-27.

[64] Larsson, M., Nagi, S. S. (2022). Role of C-tactile fibers in pain modulation: animal and human perspectives. *Current Opinion in Behavioral Sciences,* 43, 138-144

[65] Levin, J., "Functionalism", The Stanford Encyclopedia of Philosophy (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = ¡https://plato.stanford.edu/archives/sum2023/entries/functionalism/¿.

[66] Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences,* 35(3), 121-143.

[67] Lucas, T. (1961) Minds, Machines and Gödel. *Philosophy*, 36 (137) 112-127

[68] MacCorquodale, K. (1970) On Chomsky's review of Skinner's Verbal Behavior. *Journal of the experimental Analysis of Behavior*, 13, 83-99

[69] Maimon, A., Hemmo, M. (2022). Does neuroplasticity support the hypothesis of multiple realizability?. *Philosophy of Science,* 89(1), 107-127.

[70] Makin, T. R., & Krakauer, J. W. (2023). Against cortical reorganisation. *elife,* 12, e84716.

[71] Marr, D., Hildreth (1980) Theory of Edge Detection. *Proc. Roy Soc .Lon. B*, 207, 187 – 217

[72] Marr, S. Vision. A computational investigation into the the human representation and proceessing of visual information. New York 1982

[73] McCulloch, W., Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics,* 5, 115–133

[74] Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in neurosciences,* 6, 414-417

[75] Murray J.D.: Mathematical Biology. Heidelberg 1989

[76] Nagel, E.: The meaning of reduction in the natural sciences. In: Danto, A., Morgenbesser, S., (eds), Philosophy of Science, Cleveland 1960, pp. 288–312

[77] Nagel, E.: The structure of science. New York 1961

[78] Nagel, Thomas (1974) What Is It Like to Be a Bat? *The Philosophical Review,* 83 (4): 435–450

[79] Nelson, F. (2023) Our brains can't actually 'rewire' themselves, neuroscientis say. *sciencealert*, November 2023, https://www.sciencealert.com/the-brain-is-not-able-to-rewire-itself-neuroscientists-say

[80] Ney, Reductionism. *Internet Encyclopedia of Philosophy,* 2006

[81] Nisbett, R., Wilson, T. (1977), Telling More Than We Can Know: Verbal Reports on Mental Processes, *Psychological Review* 84, 231-259

[82] Noë, A.: Du bist nicht dein Gehirn – Eine radikale Philosophie des Bewußtseins. München 2010

[83] Northoff, G., Heinzel, A. (2006) First-person neuroscience: a new methodological approach for linking mental and neuronal states. *Philosophy, Ethics, and Humanities in Medicine,* 1, 1-10.

[84] Pascual-Leone, A., Hamilton, R. (2001) The Metamodal Organization of the Brain. *Progress in Brain Research,* vol. 134, eds. C. Casanova and M. Ptito, 427–45. Elsevier Science

[85] Piccinini, G. (2004) Functionalism, Computationalism, and Mental States. *Stud. Hist. Phil. Sci*, 35, 811 - 853

[86] Piccinini, G. (2006). Computational explanation in neuroscience. *Synthese,* 153, 343-353.

[87] Piccinini, G., Shagrir, O. (2014). Foundations of computational neuroscience. *Current Opinion in Neurobiology,* 25, 25–30. https://doi.org/10.1016/J.CONB.2013.10.005

[88] Piccinini, G., Shagrir, O. (2014). Foundations of computational neuroscience. *Current Opinion in Neurobiology,* 25, 25-30.

[89] Pinna, S. (2011). The Turing Machine as a Cognitive Model of Human Computation. In: Le scienze cognitive in Italia 2011. AISC'11 (pp. 147-149). Universita degli Studi di Napoli "Federico II"

[90] Place, U.T. (1956) Is Consciousness a Brain Process?, *British Journal of Psychology,* 47: 44–50

[91] Polger, T.W. Functionalism. *Internet Encyclopedia of Philosophy (IEP)*

[92] Popper, K.R. (1957) The aim of Science, *Ratio*, I, 24–35

[93] Popper, K.R.: Truth, Rationality, and the Growth of Scientific Knowledge, in: Popper, K.R. Conjectures nd Refutations, New York 1962, 24–35

[94] Putnam, H. (1967) Psychological predicates. *Art, Mind, and Religion*, 1, 37-48.

[95] Putnam, H.: Nature of Mental States, Psychophysical Predicates, und The mental life of some machines. In H. Castaneda (Ed.) Intentionality, minds, and perception (pp. 177 – 200), Detroit 1967

[96] Putnam, H. 1967b: "The Nature of Mental States" (originally published as "Psychological Predicates"), in Captain, W. H. and Merrill, D. D. (eds.), Art, Mind and Religion, Pittsburgh: University of Pittsburgh Press, pp. 37-48. Reprinted in Putnam, H. 1975a: pp. 429-440.

[97] Putnam, H. Minds and Machines. In: Putnam, H. (ed), Mind, Language and Reality: Philosophical Papers. Vol. 2. New York: Cambridge University Press; 1975

[98] Putnam, H. The Meaning of "Meaning". In: Putnam, H.: Mind, Language and Reality, Philosophical Papers, Volume 2. Cambridge University Press 1975

[99] Putnam, H.: Representation and Reality. Cambridge (Mass.), London 1988, pp. 121 – 125

[100] Putnam, H.: The Nature of Mental States, Psychophysical Predicates, und The mental life of some machines. In H. Castaneda (Ed.) Intentionality, minds, and perception (pp. 177 – 200), Detroit 1967

[101] Putnam, H. Functionalism: cognitive science or science fiction? (pp. 32-44). Oxford University Press 1997

[102] Putnam, H. *Philosophy of Mind: The Key Thinkers,* 147. Bloomsbury Academic 2013

[103] Quine, W.V. (1964) Ontological Reduction and the World of Numbers. *Journal of Philosophy*, LXI, March 26, 209–216

[104] Rabinowitz, N. C., Willmore, B. D., Schnupp, J. W., & King, A. J. (2011). Contrast gain control in auditory cortex. *Neuron,* 70(6), 1178-1191.

[105] Richardson, R. (1982) How not to reduce a functional psychology. *Philosophy of Science*, 49, 125–137

[106] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation, parallel distributed processing, Explorations in the microstructure of cognition, ed. D.E. Rumelhart and J. McClelland. vol. 1. 1986. *Biometrika,* 71(599-607), 6.

[107] Sadato, N., Pascual-Leone, A., Grafman, J., Ibañez, V., Deiber, M. P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature,* 380(6574), 526-528.

[108] Sadato, N. (2005a), "How the Blind "See" Braille: Lessons from Functional Magnetic Resonance Imaging." *The Neuroscientist* 11, 577–82.

[109] Sadato, N., Okada, T., Honda, M., Matsuki, K. I., Yoshida, M., Kashikura, K. I., ..., & Yonekura, Y. (2005b). Cross-modal integration and plastic changes revealed by lip movement, random-dot motion and sign languages in the hearing and deaf. *Cerebral Cortex,* 15(8), 1113-1122.

[110] Saur, D., Lange, R., Baumgaertner, A., Schraknepper, V., Willmes, K., Rijntjes, M., & Weiller, C. (2006). Dynamics of language reorganization after stroke. *Brain,* 129(6), 1371-1384.

[111] Scheibe, E.: Die Philosophie der Physiker. München 2006

[112] Schleichert,H.(Hg.), Logischer Empirismus - der Wiener Kreis, München: 1975,201-222

[113] Schwarz, W. (2013). Contingent identity. *Philosophy Compass*, 8(5), 486-495.

[114] Searle, J.: The rediscovery of the mind. Cambridge MA, MIT 1992

[115] Sharma, J., Angelucci, A., Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature,* 404(6780), 841-847.

[116] Shagrir, O. (2005). The rise and fall of computational functionalism. *Hilary Putnam,*, 1, 220-250.

[117] Shagrir, O. (2006). Gödel on Turing on computability. *Church's Thesis after,* 70, 393-419.

[118] Shagrir, O.:Rise and fall of computational functionalism. In: Y. Ben-Menachem (ed) Hilary Putnam. Cambridge University Press 2005, p. 220–250)

[119] Shagrir, O. (2014). Hilary Putnam and computational functionalism. *Philosophy of Mind: The Key Thinkers,* 147.

[120] Shagrir, O. (1998) Multiple Realization, Computation and the Taxonomy of Psychological States, *Synthese*, 114 (3), 445-461

[121] Schaffner, K. F. (1967). Approaches to reduction. *Philosophy of Science*, 34(2), 137-147

[122] Shapiro, L. A. (2008). How to test for multiple realization. *Philosophy of Science,* 75(5), 514-525

[123] Shapiro, L. A., & Polger, T. W. (2012). Identity, variability, and multiple realization in the special sciences. *New perspectives on type identity: The mental and the physical,* 264-288.

[124] Shapiro, L. A. (2000). Multiple realizations. *The Journal of Philosophy,* 97(12), 635-654

[125] Shapiro, L. A., & Polger, T. W. (2012). Identity, variability, and multiple realization in the special sciences. New perspectives on type identity: The mental and the physical, 264-288

[126] Sharma, J., Angelucci, A., & Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature,* 404(6780), 841-847

[127] Singer, W. (2007) Binding by synchrony, *Scholarpedia,* 2(12):1657.

[128] Siegelmann, H. T. (1995). Computation beyond the Turing limit. *Science,* 268(5210), 545-548.

[129] Shlizerman, E. Schroeder, K. Kutz (2012)Neural activity measures and their dynamics,, *SIAM J. APPL. MATH. Society for Industrial and Applied Mathematics*, Vol. 72, 4, pp. 1260–1291

[130] Slezak, P. (1984). Minds, Machines and Self-Reference. *Dialectica,* 38(1), 17-34.

[131] Smart, J.J.C. (1959) Sensations and Brain Processes, *Philosophical Review,* 68, 141–156.

[132] Sprevak, M.: Turing's model of the mind. In: Copeland, J., Bowen, M., Sprevak,M., Wilson, R.(eds) *The Turing Guide: Life, Work, Legacy,* Oxford UniversityPress 2017

[133] Suppe, F.: The Semantic Conception of Theories and Scientific Realism. Urbana: University of Illinois Press 1989

[134] Suppes, P.: Introduction to Logic. Princeton: D. Van Norstrand Co. Inc. 1957

[135] Sur, M., Angelucci, A., Sharma, J. (1999). Rewiring cortex: The role of patterned activity in development and plasticity of neocortical circuits. *Journal of neurobiology,* 41(1), 33-43.

[136] Strogatz, Nonlinear dynamics and chaos. Cambridge University Press 1994

[137] Tanaka, A., Tomiya, A., Hashimoto, K. (2021). Deep learning and physics (Vol. 1). Singapore: Springer

[138] Tian, T., Zhu, Y. L., Hu, F. H., Wang, Y. Y., Huang, N. P., & Xiao, Z. D. (2013). Dynamics of exosome internalization and trafficking. *Journal of cellular physiology,* 228(7), 1487-1495.

[139] Turing, A. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society Series 2* (42), 230-42.

[140] Turing, A. M. (1950). Computing Machinery and Intelligence *Mind*, 59(236), 433-460.

[141] Turing, A. M. (1950). Mind. *Mind,* 59(236), 433-460.

[142] van Gelder, T. (1995) What might cognition be, if not computation? *The Journal of Philosophy* , 92(7), 345 – 381

[143] van Gelder, T. (1998) The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21, 615–665

[144] von Melchner, L., Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature,* 404(6780), 871-876.

[145] von Neumann, J. 1951. The General and Logical Theory of Automata. *Cerebral Mechanisms in Behavior.* L. A. Jeffress. New York, Wiley: 1–41.

[146] Wason, P. C. (1968), Reasoning about a rule. *Quarterly Journal of Experimental Psychology,* 20 (3): 273-281

[147] Watson, J. B. (1913) Psychology as the Behaviorist Views It. In: *Psychological Review,* 20, 158–177

[148] Wells, A. (1998) Turing's Analysis of Computation and Theories of Cognitive Architecture, *Cognitive Science*, 22:269-294

[149] Woodger, H. Biology and Language, Cambridge, Cambridge University Press 1952

[150] Worsch, T. Modelle der Parallelverarbeitung, Teil 1 Turingmaschinen über parallelverarbeitende TMs, ppt. Institut für Theoretische Informatik, Karlsruher Institut für Technologie, Sommersemester 2020

[151] Zangwill, N. (1992), Variable Realization: Not Proved, *The Philosophical Quarterly*, 42(167): 214–219. doi:10.2307/2220216

[152] Zapparoli, L., Seghezzi, S., Paulesu, R. (2017) The What, the When, and the Whether of Intentional Action in the Brain: A Meta-Analytical Review. *Frontiers in Human Neuroscience*, 11:238, doi: 10.3389/fnhum.2017.00238

Index