

Neurokognitive Modellierung

Neurocognitive modeling

Arthur M. Jacobs, Markus J. Hofmann

Freie Universität Berlin
Allgemeine und Neurokognitive Psychologie
Dahlem Institute for Neuroimaging of Emotion (D.I.N.E.)
Habelschwerdter Allee 45
14195 Berlin
email: ajacobs@zedat.fu-berlin.de

0 Zusammenfassung

Neurokognitive Modelle sollen helfen, Hypothesen über die Rolle der Hirnaktivität beim Funktionieren mentaler Vorgänge zu spezifizieren, so dass sie mittels neurokognitiver Methoden geprüft und falsifiziert werden können. Praktisch bedeutet dies, zu spezifizieren unter welchen Bedingungen ein Reizkontext welche behavioralen und neuronalen Reaktionen auslöst und wie diese mit subjektivem Erleben in Zusammenhang stehen. Da neurokognitive Methoden eine Vielzahl neuer abhängiger Variablen (AVs) liefern, stehen neurokognitive Modelle vor der Herausforderung, neben Verhaltensdaten auch diese neuen AVs qualitativ oder quantitativ im Rahmen eines psychologischen Experiments vorherzusagen. Wir diskutieren Probleme, Lösungsstrategien und Evaluationskriterien für dieses Unterfangen.

1 Einleitung

Eine Suche mit dem Stichwort „neurocognitive model“ in Science Direct ergab im Februar 2011 mehr als 600 Antworten! Solche neurokognitiven Modelle sind heute in praktisch alle Bereiche der Psychologie eingedrungen und werden längst nicht mehr nur in den klassischen theoriebildenden Zeitschriften des Fachs wie Psychological Review oder Cognitive Science, sondern noch häufiger in Journalen wie Neurocomputing, Neural Networks, Nature Reviews Neuroscience, Trends in Cognitive Sciences usw. publiziert. Unter den Antworten der Suchmaschine befinden sich zahlreiche Modelle zu kognitiven oder affektiven Prozesse (Perzeption, Aufmerksamkeit, Gedächtnis, Sprache, Belohnungslernen, Emotionsentstehung), aber auch zu klinischen Störungen (Depression, Schlafstörung, Alkoholismus, Schizophrenie) oder Anwendungen (Navigation, Robotik, Neuroprothetik).

Was sollen und können diese Modelle leisten? Die Psychologie kann heute – in Anspielung auf die klassische Definition – als die „Wissenschaft vom Erleben, Verhalten und der diesen

zugrundeliegenden Hirnaktivität“ bezeichnet werden. Wer heute Psychologie studiert, muss in praktisch jedem Bereich des Fachs lernen, „subjektives“ Erleben, „objektives“ (intersubjektiv prüfbares) Verhalten und Hirnaktivität unter einen Hut zu bringen. Auch zu Zeiten Wolfgang Köhlers wurde dies natürlich bereits versucht, aber erst die modernen bildgebenden Verfahren haben mit ihrer rasanten Entwicklung seit Beginn der 90er Jahre dazu beigetragen, dass die „abhängige Variable Hirnaktivität“ quasi flächendeckend die Fragestellungen und Befundinterpretationen des Faches bestimmt. Dies verändert auch die Beschreibungs- und Erklärungsstrategien und damit die Modellierungspraktiken der Psychologie.

Ganz allgemein sollen Neurokognitive Modelle helfen, Hypothesen über die Rolle der Hirnaktivität beim Funktionieren mentaler Vorgänge zu spezifizieren, so dass sie mittels neurokognitiver Methoden wie fMRT, EEG, TMS oder fNIRS geprüft und falsifiziert werden können (Jacobs, Hutzler, & Engl, 2005). Praktisch bedeutet dies, zu spezifizieren welche behavioralen und neuronalen Reaktionen ein Reizkontext unter welchen Bedingungen auslöst und wie diese mit subjektivem Erleben in Zusammenhang stehen. Da neurokognitive Methoden eine Vielzahl neuer AVs liefern, stehen neurokognitive Modelle vor der Herausforderung, neben Verhaltensdaten auch diese neuen AVs qualitativ oder quantitativ im Rahmen eines psychologischen Experiments vorherzusagen.

Bezüglich der psychologischen Modellbildung unterscheiden sich Methoden der Hirnaktivitätsmessung hinsichtlich der Frage, ob die Anzahl räumlicher oder zeitlicher Messpunkte vorrangig ist, d.h. ob sie primär Hypothesen zur Beantwortung der Wo- oder der Wann-Frage prüfen helfen (Jacobs, 2006). Aber natürlich implizieren Antworten auf Wo- oder Wann-Fragen noch keine Klärung der Wie-Frage nach dem Funktionieren mentaler Vorgänge (Aue, Lavell, & Cacioppo, 2009; Schrott & Jacobs, 2011; Lorig, 2009). Wie spezielle mentale Prozesse des Denkens, Fühlens oder Sprechens unter Einbeziehung ihrer vermeintlichen neuronalen Substrate funktionieren, diese Frage kann nicht allein durch den Einsatz neurokognitiver Methoden beantwortet werden; dazu benötigt man neurokognitive Modelle.

Einschränkend wollen wir im Folgenden den Fokus auf psychologische neurokognitive Modelle legen, die bereits geprüfte Vorhersagen über Hirnaktivitätsdaten in psychologischen Experimenten erlauben. Dies trifft auf das Gros der in Zeitschriften wie Neurocomputing, Neural Networks oder Frontiers in Neuroinformatics publizierten Modelle nicht zu, deren Diskussion auch den Rahmen dieses Aufsatzes sprengen würde.

2 Vom Flussdiagramm zur Vorhersage hirnelektrischer und hämodynamischer Antwortfunktionen: ein langer Weg!

Die klassischen Flussdiagramm-Modelle der Neurologie und Psychologie (Kästchen- und Pfeil Diagramme, „Boxologische Modelle“; für eine Klassifikation s. Jacobs & Grainger, 1994), die von den historischen „diagram makers“ Wernicke und Lichtheim über Broadbents (1958) „boxologisches“ Standardmodell der Informationsverarbeitung bis hin zu aktuellen Modellen der Sprachverarbeitung oder des Lesens reichen (Jacobs, 2003, 2008, 2011) dienen in der Regel dazu, Hypothesen über das Zustandekommen von Verhaltensdaten wie Fehlerraten oder Reaktionszeiten zu spezifizieren. Dies gilt ebenso für die beiden anderen Standard-Modellformate der Psychologie (mathematische und Simulationsmodelle; Jacobs, 2008).

Der Prototyp aller konnektionistischen Simulationsmodelle, das Interaktive Aktivationsmodell (McClelland & Rumelhart, 1981), beispielsweise leistete Vorhersagen über Fehlerraten in einer spezifischen Aufgabe, dem sog. Reicher Paradigma, um einen spezifischen Effekt zu erklären, den sog. Wortüberlegenheitseffekt. Erst in den 90er Jahren wurden darauf aufbauende Simulationsmodelle, die auch Reaktionszeiten und Reaktionszeitverteilungen in anderen Aufgaben als dem Reicher-Paradigma vorhersagen konnten, entwickelt (Jacobs & Grainger, 1994; Grainger & Jacobs, 1996). Jedoch sind auch dies lediglich Maße des behavioralen Endprodukts, der Gesamtverarbeitungszeit, die ein bestimmter mentaler Vorgang benötigt, um eine aufgabenadäquate Reaktion zu produzieren; die Prozessdynamik wird nicht direkt erfasst und modelliert (Barber & Kutas, 2007; Jacobs, 2006). Auch sog. „biologisch plausible“ Simulationsmodelle, wie etwa das vielzierte „Complementary Learning Systems“ Modell (z.B. O'Reilly & Norman, 2002) erlauben in der Regel „nur“ die Vorhersage von Verhaltensdaten.

1995 spekulierten Jacobs und Carr, dass Simulationsmodelle die Kognitive Neurowissenschaft dadurch befruchten könnten, dass sie prüfbare Vorhersagen über Hirnaktivierungsdaten in bestimmten Aufgaben leisten, etwa zu N400 Amplituden oder hämodynamischen Antwortfunktionen. Vier Jahre später stellten Jacobs und Rösler (1999) immer noch keinen nennenswerten Fortschritt in der 1995 geforderten Kreuzbefruchtung zwischen neurokognitiven Modellierern und „mind mappers“ fest. In einem rund 10 Jahre später erschienenen Übersichtsartikel über die Rolle von Simulationsmodellen in der elektrophysiologischen Leseforschung – einem seit Jahrzehnten hochintensiv beforschten Gebiet – zitieren Barber und Kutas (2007) auch lediglich zwei klare Beispiele für ein solches Vorgehen: die Arbeiten von Holcomb, Grainger und O'Rourke (2002) und Braun, Jacobs, Hahne, Ricker, Hofmann und Hutzler (2006). Heute finden sich immerhin einige weitere Arbeiten, die einen quantitativen Zusammenhang zwischen Modell- und EKP-

Aktivierungen bei der Worterkennung herstellen (Hofmann, Tamm, Braun, Dambacher, Hahne, & Jacobs, 2008; Huber, Tian, Curran, O'Reilly, & Worocho, 2008; Klonek, Tamm, Hofmann, & Jacobs, 2009; Rey, Dufau, Massol, & Grainger, 2009). Auch außerhalb der Leseforschung kommen solche Modelle erst recht spärlich zum Einsatz, um neurokognitive Befunde zu interpretieren, wie ein Blick in aktuelle Ausgaben von Zeitschriften wie *Psychophysiology*, *Biological Psychology*, *Journal of Cognitive Neuroscience* oder *Brain Research* offenbart. Ein interessantes Beispiel dafür, wie Simulationsmodelle die neurokognitive Forschung befruchten können, stellt sicher der Theoriestreit um die Funktionsweise des anterioren cingulären Cortex dar, der durch konkurrierende Simulationsmodelle von Botvinick, Braver, Barch, Carter, und Cohen (2001), Holroyd und Coles (2002), oder Brown und Braver (2005), entfacht wurde.

Ein exemplarisches, deskriptiv-, „boxologisches“ neurokognitives Modell der auditiven Satzverarbeitung, das primär auf Befunden aus EKP Studien aufbaut, stammt von Friederici (2002). Das Modell verbindet kognitive Prozesse wie Phonemidentifikation oder Integration von semantischer und morphosyntaktischer Information sowohl mit EKP Komponenten wie N100 oder N400, als auch mit Hirnregionen wie BA44 oder MTL. Es spezifiziert nicht, wie diese einzelnen kognitiven Subprozesse funktionieren, ist also kein Prozessmodell.

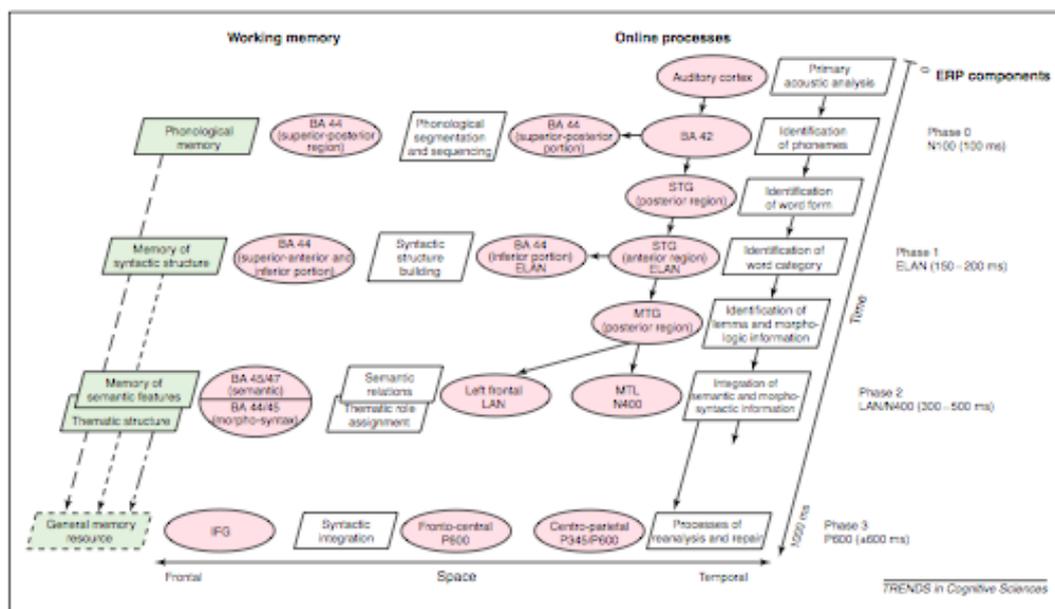


Abb. 1. Neurokognitives Modell der auditiven Satzverarbeitung mit hypothetischen Beziehungen zwischen kognitiven Vorgängen, EKP Komponenten und Hirnregionen (Friederici, 2002).

Auch bezüglich der zweiten weitverbreiteten neurokognitiven Methode, fMRT, kann aktuell keine Rede davon sein, dass neurokognitive Modelle –egal ob vom „boxologischen“, vom

mathematischen oder vom komputationalen Typ- standardmäßig zur Vorhersage oder Interpretation der Befunde herangezogen werden. In 2003/2004 publizierte die Gruppe um John R. Anderson eine Reihe von Artikeln in verschiedenen Zeitschriften, in denen einem größeren Leserkreis aus der Psychologie vermutlich zum ersten Mal die Vorhersage von hämodynamischen Antwortfunktionen durch ein klassisches Informationsverarbeitungsmodell aus der Familie der Produktionssysteme, dem ACT-R (Adaptive Control of Thought-rational, Anderson, 1990), demonstriert wurde (Anderson, Bothell, Byrne, Douglas, Lebiere, Quin, 2004; Anderson, Quin, Sohn, Stenger, & Carter, 2003; Anderson, Quin, Stenger, & Carter, 2004). Ähnliches hatten beispielsweise Just, Carpenter und Varma (1999) bereits mit einem weniger bekannten Produktionssystemmodell („4CAPS“) versucht. Das ACT-R modelliert die Aktivität der einzelnen Boxen („Verarbeitungspuffer“), welche wiederum über eine entsprechende Funktion die hämodynamischen Antwortfunktionen vorhersagt. Es spezifiziert auch die in den Boxen genannten kognitiven Prozesse anhand von Algorithmen, nicht aber das Funktionieren der entsprechenden neuronalen Prozesse. In diesem Sinne kann es als partielles neurokognitives Prozessmodell gelten.

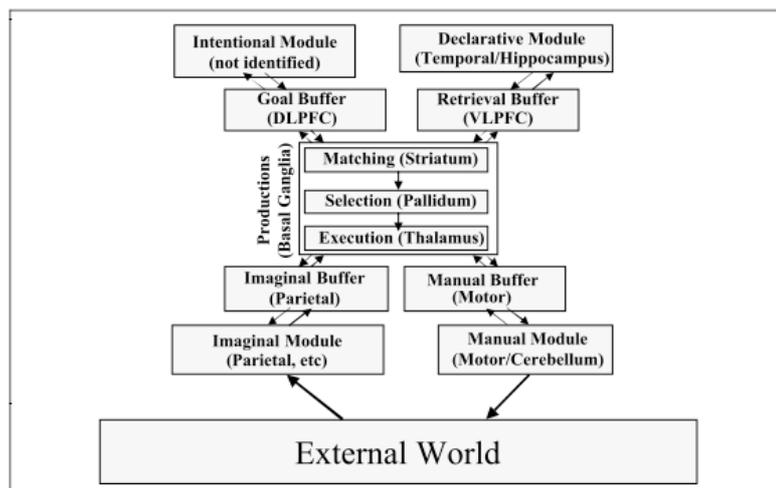


Abb. 2. Skizze des ACT-R 5.0 mit hypothetischen Beziehungen zwischen kognitiven Vorgängen (z.B. Abgleich/matching, Ausführung/execution) und Hirnregionen (z.B. Striatum, Thalamus).

3 Gehirn und Geist: eine komplexe Beziehungskiste mit umgekehrter Inferenz

Ein grundsätzliches Problem dieser Modellierungsansätze betrifft die übergeordnete Frage nach der Beziehung zwischen kognitiven (oder affektiven) und neuronalen Prozessen, unabhängig davon, ob diese mittels EEG oder fMRT analysiert werden. Wenn etwa das ACT-R Modell dazu benutzt wird, hämodynamische Antwortfunktionen in parietalen und präfrontalen Hirnregionen in einer Algebra Aufgabe vorherzusagen, oder das Modell von Friederici eine Beziehung zwischen einem kognitiven Prozess und einer bestimmten EKP

Komponente hypostasiert, stellt sich die entscheidende Frage, ob diese Hirnregionen bzw. EKP Komponenten eine notwendige und/oder hinreichende, eine mitwirkende oder schlichtweg nur eine inzidentelle Rolle bei den kognitiven Vorgängen oder Zuständen spielen, welche die Modelle hypostasieren. Träfe letzteres systematisch zu, wäre der Einsatz der relativ kostspieligen neurokognitiven Methoden zur Erklärung bestimmter, in psychologischen Experimenten beobachteter Verhaltenseffekte zu hinterfragen. Wie Price und Friston (2005) ausführen, besteht diese Möglichkeit aber in praktisch jedem fMRT Versuch.

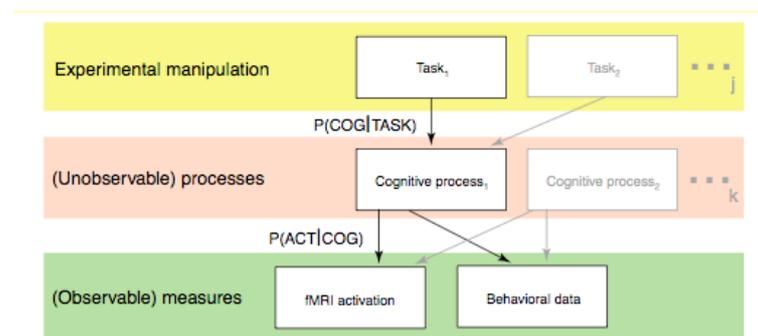


Abb. 3. Probabilistischer Graph (Bayes Netz), der die Beziehungen zwischen experimentellen Manipulationen (unabhängigen Variablen/UVs), kognitiven Prozessen und AVs repräsentiert. Die Pfeile repräsentieren bedingte Auftretenswahrscheinlichkeiten zwischen den Entitäten in den Kästchen (Poldrack, 2006).

Das in Abbildung 3 skizzierte Bayes Netz verdeutlicht das Problem der „many-to-many“ Abbildungen zwischen kognitiven und neuronalen Vorgängen. Der probabilistische Graph macht einige Merkmale der umgekehrten Inferenz deutlich. Umgekehrte Inferenz ist der (logisch invalide) Schluss von der Konsequenz auf die Antezedenz („Bejahung der Konsequenz“), aber als Abduktion nützlich für die Entdeckung neuer Fakten (Poldrack, 2006). Beispielsweise kann man aus der Regel „Wo Feuer ist, das ist auch Rauch“ vom Vorhandensein des Rauchs nicht eindeutig auf das Vorhandensein von Feuer schließen (es sei denn, Feuer wäre als notwendige und hinreichende Bedingung, d.h. alleinige Ursache, für Rauch etabliert). In der Praxis ist die Abduktion „Wo Rauch ist, da ist (wahrscheinlich) auch Feuer“ jedoch eine nützliche Heuristik. In zahlreichen Artikeln aus den kognitiven Neurowissenschaften wird ähnlich „abduziert“, wenn man beispielsweise aus der wiederholten Beobachtung „in Angstinduktionstests glüht die Amygdala“ schließt: „in unserem Test glühte die Amygdala, also lag Angst vor“. Da unwahrscheinlich ist, dass diese Art der umgekehrten Inferenz aus neurowissenschaftlichen Studien einfach verbannt werden kann und ihr Nutzen für die Generierung neuer Hypothesen immer wieder

herausgestellt wird, schlagen prominente Vertreter dieser Forschungsrichtung wie Friston, Price oder Poldrack vor, solche Abduktionen sehr vorsichtig einzusetzen, vor allem wenn die beiden in Abb. 3 skizzierten bedingten Wahrscheinlichkeiten, i) die Selektivität der Aktivierung des betroffenen Hirnareals [$p(\text{ACT}/\text{COG})$] und ii) der Glaube an das tatsächliche Ablaufen eines bestimmten kognitiven Prozesses [$p(\text{COG}/\text{TASK})$], gering sind. Jedoch kann an der Maximierung dieser bedingten Wahrscheinlichkeiten gearbeitet werden, wobei neurokognitive Modelle eine entscheidende Rolle spielen (Lenartowicz, Congdon, & Poldrack, 2010; Poldrack, Halchenko, & Hanson, 2009).

4 Wie baue ich ein gutes neurokognitives Modell? Funktionale Ontologien und Modellevaluationskriterien

1996 veröffentlichte David Poeppel seine wegweisende Kritik der damaligen PET Studien zu der Frage, wo im Gehirn „Phonologie“ verarbeitet wird. Diese Kritik birgt auch einige Lehren für neurokognitive Modellierer. Mit Bezug auf die Maximierung von $p(\text{COG}/\text{TASK})$ ist natürlich der Begriff „Phonologie“ unterspezifiziert und jedes Modell der phonologischen Verarbeitung sollte möglichst genau spezifizieren, welche (hypothetischen) phonologischen (Sub-)Prozesse durch die verwendete Aufgabe aktiviert werden, analog zu Sternbergs (1969) in dieser Hinsicht wegweisendem Modell des Zugriffs auf das verbale Kurzzeitgedächtnis.

Je genauer die Prozessbeschreibung bzw. Implementierung in einem Simulationsmodell im Rahmen der verwendeten Aufgabe ist, desto höher sollte $p(\text{COG}/\text{TASK})$ sein. Jacobs und Grainger (1994) schlugen zur Erhöhung der Spezifität von (neuro)-kognitiven Modellen mehrere Strategien vor. Hier sei nur auf die Strategie des „functional overlap modeling“ verwiesen, deren Ziel die Identifizierung der Kernprozesse ist, die der Leistung in *allen* experimentellen Paradigmen unterliegen, welche vermeintlich den gleichen kognitiven Vorgang, z.B. Worterkennung, rekrutieren. Im Bereich der Leseforschung sind dies beispielsweise die lexikalische Entscheidungsaufgabe, die perzeptive Identifikationsaufgabe oder die Benennungsaufgabe. All diese Tests rekrutieren sowohl unterschiedliche als auch gemeinsame kognitive Vorgänge, die ein gutes Modell explizit und damit prüfbar machen sollte. Die Worterkennungsforschung bietet eine Reihe von Beispielen für die erfolgreiche Anwendung dieser Strategie auf die Konstruktion von kognitiven Simulationsmodellen (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; Perry, Ziegler, & Zorzi, 2007).

Das zweite Standbein einer erfolversprechenden neurokognitiven Modellierung betrifft die Maximierung von $p(\text{ACT}/\text{COG})$. Diese hängt von zahlreichen Faktoren ab, die nicht notwendigerweise „in der Hand“ des Modellierers liegen. So beeinflusst beispielsweise die

geschätzte Größe eines Hirnareals diese bedingte Wahrscheinlichkeit. Allein aufgrund seiner Größe eignet sich die sog. Pars triangularis, die oft mit dem Broca Areal identifiziert wird und in vielen neurokognitiven Modellen eine Rolle spielt (s. Abb. 1), für umgekehrte Inferenzen eher weniger. Laut Poldrack (2006) umfasst dieses Areal einen Würfel mit Kantenlänge von rund 28 mm. Poldrack weist auf eine negative nonlineare Beziehung zwischen der Größe einer „region of interest“ (ROI) und der Selektivität ihrer neuronalen Antwort hin. Trotz informationstechnischer Hilfsmittel wie „BrainMap“ oder „PubBrain“ gestaltet sich die Schätzung von $p(\text{ACT}/\text{COG})$ schwierig, was auch darin liegt, dass diese Datenbanken Taxonomien kognitiver Prozesse verwenden, die zu grob sind. „Semantik“ oder „Arbeitsgedächtnis“ sind Beispiele für solche unterspezifizierten Begriffe, die in diesen Datenbanken weiter ausdifferenziert werden müssen, um die neurokognitive Modellierung verbessern zu helfen. Poldrack fordert in Anlehnung an Price und Friston (2005), dass die „kognitive Ontologie“ dieser Datenbanken optimiert wird (s. a. Lenartowicz et al., 2010). Price und Friston (2005) bezeichnen damit systematische Form-Funktions Beziehungen, die sich aus einer neuroanatomischen und einer kognitiven Menge speisen. Systematische Definitionen oder Modelle sollen Schlüsse erlauben von neuroanatomischen Strukturen auf kognitive Funktionen und umgekehrt. Hierbei ist entscheidend, dass das Modell die *Rolle*, die ein bestimmtes Hirnareal bei einer –möglichst präzise spezifizierten- kognitiven Funktion spielt, ebenso präzise spezifiziert. Labels wie „Visual Word Form Area“ sind laut Price und Friston dabei wenig weiterführend, da damit nur indiziert wird, dass diese Area eine Rolle bei der Funktion „Worterkennung“ spielt, aber nicht welche – erst der Bezug zu neurokognitiven Modellen schreibt solchen empirischen Begriffen eine funktionale Rolle zu (Hofmann, Kuchinke, Tamm, Vö, & Jacobs, 2009).

Die bislang erörterten Modelle beruhen meist auf modularen Grundannahmen – das heißt eine räumlich zusammenhängende Region wird im Hinblick auf ihre „Funktion“ untersucht (z. B. Anderson, 1990). Die in der ersten Dekade dieses Jahrtausends am häufigsten zitierten Autoren der Zeitschrift Nature Reviews Neuroscience halten jedoch die Netzwerk-Perspektive auf das Gehirn für einen entscheidenden Durchbruch der empirischen Forschung (Luo et al., 2010). Deshalb werden funktionale Konnektivitäts- bzw. Interaktionsstudien zukünftig eine zentrale Rolle bei der Optimierung zeitgemäßer kognitiver Ontologien spielen (vgl. Jacobs, 2006).

Price und Friston definieren zwei scharfe Kriterien für „gute“ Ontologien (s. Abb. 4):

- sie besitzen eine hierarchische Struktur, welche die Koaktivierung neuroanatomischer Regionen vorhersagt und in der Mengen von koaktivierten Arealen demonstrierbare (effektive) Verbindungen aufweisen
- sie erlauben, ausgehend von einer Aktivungsverteilung, Vorhersagen über kognitive Verarbeitung, die darauf beruhen, welches Areal bzw.

welche Kombination von Regionen notwendig für diesen Verarbeitungsvorgang ist.

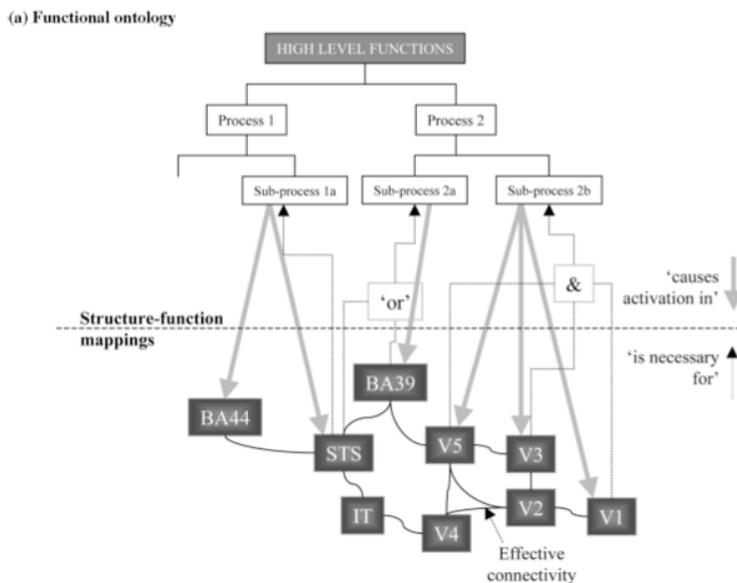


Abb. 4. Funktionale Ontologie (aus Price und Friston, 2005; Erläuterungen s. Text).

Neurokognitive Modelle sollten aber nicht allein nach den von Price und Friston (2005) vorgeschlagenen Kriterien beurteilt werden, sondern genauso wie kognitive Modelle auch die üblichen Gütekriterien psychologischer Modellbildung erfüllen. Eine ausführliche Diskussion solcher Kriterien findet sich in Jacobs und Grainger (1994, s.a. Grainger & Jacobs, 1998). Wir können hier nur eine kurze Auflistung mit einigen Beispielen bringen.

5 Modellbewertungskriterien

„Gib mir genügend freie Parameter, und ich modelliere Dir eine rosa Kuh im Schlafanzug beim Polka-Tanzen“ (unbekannter Modellierer).

a) Deskriptive Adäquatheit

Wie genau beschreibt ein Modell den zu erklärenden Datensatz? Die Frage lässt sich bei der Bewertung mathematischer oder algorithmischer Modelle, die beispielsweise eine ROC-Kurve aus einem Signalentdeckungsexperiment vorhersagen, mit statistischen Standardmethoden der Anpassungsgüte (R^2 , RMSD, χ^2 , AIC, BIC usw.) beantworten (Jacobs, Graf, & Kinder, 2003). Jedoch erlauben nur wenige neurokognitive Modelle ein ähnliches Vorgehen in Bezug auf AV wie hämodynamische Antwortfunktionen oder N400 Amplituden.

Ein für die Psychologie wegweisendes Beispiel geben Anderson et al. (2004). Sie

nehmen an, dass die Intensität der hämodynamischen Antwort mit der Dauer ansteigt, die ein Reiz in einem „Verarbeitungspuffer“ verbringt (Abb. 2). Wenn der Stimulus also längere Zeit im „retrieval buffer“ gehalten werden muss, so sagt das ACT-R eine stärkere hämodynamische Antwort im ventrolateralen präfrontalen Cortex (VLPFC) voraus (Danker, Gunn, & Anderson, 2008), die sich mit den empirisch ermittelten „BOLD-responses“ für jede Bedingung vergleichen lässt. Ein Beispiel für EKP-Amplituden liefern Hofmann et al. (2008). Sie berechnen das Ausmaß lexikalischer Kompetition, den jeder Buchstabenreiz in einem simulierten mentalen Lexikon auslöst (Botvinick et al., 2001; McClelland & Rumelhart, 1981). Dieses Konfliktmaß wird dann in eine direkte Beziehung zu den N400-Amplituden gesetzt, die durch dieselben Reize evoziert wurden. In beiden Fällen quantifiziert die Varianz, die der Modell-Prädiktor an den empirischen Daten aufklärt, wie gut das jeweilige Modell die Daten beschreibt.

b) Verallgemeinerbarkeit

Neurokognitive Modelle können nach ihrer *horizontalen* und *vertikalen* Verallgemeinerbarkeit beurteilt werden. Erstere bezieht sich auf die Fähigkeit eines Modells über unterschiedliche Reizmaterialien („stimulus generality“; Jacobs & Grainger, 1994), Paradigmen („task generality“) oder Reaktionsmasse („response generality“) zu generalisieren. Sagt ein Modell nur die N1 Amplitude in einem einzigen Paradigma mit spezifischen Reizen vorher oder sagt es hämodynamische Antwortfunktionen in verschiedenen ROIs und Aufgaben vorher? *Vertikale* Generalisierbarkeit hingegen bezieht sich auf die Fähigkeit eines Modells über verschiedene Skalen des modellierten Prozesses zu verallgemeinern. Beschreibt es nur das Endprodukt eines neurokognitiven Vorgangs das (statisch) asymptotische Verhalten (z.B. „average peak activations“), oder die komplette Dynamik (z.B. den kompletten zeitlichen Verlauf eines EKPs)? Beschreibt es makrostrukturelle oder mikrostrukturelle Veränderungen in den Daten? Kann es beispielsweise also „nur Mittelwerts-Vorhersagen für eine experimentelle Bedingung treffen, oder erlaubt es eine Vorhersage für jedes spezifisches Reiz-Reaktions Paar?

c) Einfachheit und Falsifizierbarkeit

"But there is no conservation law that requires that the description be as cumbersome as the object described" (Simon, 1969, p. 109).

Was ist der Preis für Modelle mit hoher deskriptiver Adäquatheit und Generalisierbarkeit? Oft ist es ihre Komplexität. Aber Kopernikus, Einstein, Simon

und viele andere Theoretiker waren sich darin einig, dass Modelle immer so einfach wie möglich sein sollten. Die Komplexität eines Modells kann beispielsweise über die Anzahl (oder Länge) der verwendeten Gleichungen und/oder freien Parameter in quantitativen Modellen geschätzt werden. Bei boxologischen Modellen sind die Anzahl der Kästchen und Pfeile aufschlussreich (Jacobs, 2008).

Ein typisches „boxologisches Modell“ des Lesens aus der klinischen Neuropsychologie, das aus neun Kästchen und 14 Pfeilen besteht, kann beispielsweise $9 + 14 = 23$ reine Störungsbilder „erklären“ (Kombinationen von Kästchen und/oder Pfeilen ausgeschlossen). Selbst wenn die Literatur der aktuellen Neuropsychologie für jeden dieser modelltheoretisch möglichen 23 Fallbeschreibungen ein Beispiel lieferte (was unwahrscheinlich ist), so stellt sich die Frage, ob nicht ein Modell, welches sämtliche logisch möglichen Fälle beinhaltet eine Tautologie darstellt und damit nicht falsifizierbar ist. Die gleiche Frage kann an das neurokognitive Modell in Abb. 2 gerichtet werden. Die Beurteilung des Modells sollte aber fairerweise alle drei bisher erläuterten Kriterien in der Synopse betrachten (Pitt, Myung, & Zhang, 2002).

e) Explanatorische Adäquatheit

Das letzte hier behandelte Kriterium ist auch das schwierigste. Wie gut erklärt ein Modell einen Befund? Diese Frage kann in der Psychologie, Biologie und den kognitiven Neurowissenschaften –unabhängig davon, welchen erklärungstheoretischen Ansatz man bevorzugt- am besten relativ beantwortet werden, d.h. im direkten Wettstreit zwischen alternativen Modellen am gleichen Datensatz („strong scientific inference“; Aue et al., 2009; Jacobs & Grainger, 1994; Massaro & Cowan, 1993).

Die wichtigste Frage dabei lautet: Wie viele ad hoc Annahmen macht das Modell (vgl. 5.3), und wie viele Phänomene erklären diese Annahmen darüber hinaus? Reine ad hoc Annahmen beschreiben nur genau das Phänomen, für das sie eingeführt wurden. Sobald die Annahme weitere Phänomene erfolgreich vorhersagen kann (vgl. 5.2), gewinnt sie an Erklärungswert. Das Modell, das die meisten Vorhersagen auf der Basis möglichst weniger Annahmen macht, gewinnt. Selbstverständlich sollte man hier nicht außen vorlassen, wie gut das Modell die neuen Daten erklärt (vgl. 5.1).

Dies lässt sich am Beispiel von Gedächtnisaufgaben verdeutlichen. Beschreibt ein Messmodell gelernte und nicht-gelernte Wörter als zwei (Normal-)Verteilungen ungleicher Varianz, so erklärt dieses Modell noch nichts, weil aus einem rein deskriptiven Modell keine empirisch prüfbareren Hypothesen abgeleitet werden können (z.B. Glanzer, Kim, Hilford, & Adams, 1999; Green & Swets, 1966; Ratcliff, Sheu, & Gronlund, 1992). Man kann jedoch die Anzahl an ad hoc Annahmen bestimmen: 1.

Gelernte Wörter weisen ein stärkeres „Gedächtnissignal“ auf als nicht-gelernte Wörter. 2. Die Signal-Verteilung für gelernte Wörter hat eine größere Varianz als die für nicht-gelernte Wörter. Implementiert man Annahme 1 in eine (neuro-)kognitive Modell-Architektur, die bereits andere Phänomene erfolgreich vorhergesagt hat (McClelland & Rumelhart, 1981; Grainger & Jacobs, 1996; vgl. Hofmann et al., 2008), so ergibt sich Annahme 2 als logische Konsequenz aus Annahme 1 (Hofmann, Kuchinke, Biemann, Tamm, & Jacobs, 2011). Folglich vergrößert dieses Ergebnis den Erklärungswert des Modells, und erlaubt Vorhersagen, die mit neurokognitiven Methoden geprüft werden können (vgl. Hofmann et al., 2008).

6 Ausblick

Zusammenfassend kann man sagen, dass die in Abschnitt 5 erläuterten Modellbewertungskriterien in typischen Publikationen der kognitiven Neurowissenschaften keine explizite Rolle spielen und jedenfalls nicht systematisch in dieser Kombination angewandt oder auch nur thematisiert werden. Das mag auch daran liegen, dass die theoretische Modellbildung und die Datenerhebung in diesem noch recht jungen, aber rapide expandierenden und methodisch äußerst aufwendigen Forschungsfeld noch nicht so zusammengewachsen sind, wie dies zumindest für einige Subdisziplinen der Psychologie gilt (Jacobs & Grainger, 1994; Jacobs & Carr, 1995). Sehr viel Energie fließt zu Recht in die Entwicklung von immer neuen Messmodellen und Datenanalysemethoden. Dennoch sind Ansätze wie die hier geschilderten, insbesondere derjenige von Price und Friston (2005), vielversprechend und lassen in Kombination mit allgemeinen Theorien zur Funktionsweise des Gehirns (Friston, 2010) auch große Fortschritte in der von Jacobs und Carr oder Barber und Kutas angedachten Richtung erwarten.

Literatur

- Anderson, J. R. (1990). The Adaptive Character of Thought. Laurence Erlbaum Associates, Hillsdale, NJ.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111, 1036-1060.
- Anderson, J. R., Qin, Y., Sohn, M.-H., Stenger, V. A., & Carter, C. S. (2003). An information-processing model of the BOLD response in symbol manipulation tasks. *Psychonomic Bulletin & Review*, 10, 241– 261.
- Anderson, J. R., Qin, Y., Stenger, V. A., & Carter, C. S. (2004). The relationship of three cortical regions to an information-processing model. *Journal of Cognitive*

- Neuroscience*, 16, 637-653.
- Aue, T., Lavelle, L. A., & Cacioppo, J. T. (2009). Great expectations: What can fMRI research tell us about psychological phenomena? *International Journal of Psychophysiology*, 73, 10–16.
- Barber, H. A., & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews*, 53, 98-123.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624-652.
- Braun, M., Jacobs, A. M., Hahne, A., Ricker, B., Hofmann, M., & Hutzler, F. (2006). Model-generated lexical activity predicts graded ERP amplitudes in lexical decision. *Brain Research*, 1073-1074, 431-439.
- Brown, J., & Braver, T. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, 307, 1118-1123.
- Broadbent, D. E. (1958). Perception and communication. London: Pergamon Press.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204–256.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78–84.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11, 127-138.
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103, 518-565.
- Grainger, J., & Jacobs, A.M. (1998). (Eds.) Localist connectionist approaches to human cognition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hofmann, M. J., Tamm, S., Braun, M. M., Dambacher, M., Hahne, A., & Jacobs, A. M. (2008). Conflict monitoring engages the mediofrontal cortex during nonword processing. *NeuroReport* 19, 25-29.
- Hofmann, M. J., Kuchinke, L., Tamm, S., Võ, M. L. H., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective, & Behavioral Neuroscience*, 9, 389-397.
- Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, 14, 938-950.
- Holroyd, C. B., Coles, M. G. (2002). The neural basis of human error processing:

- Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109, 679-709.
- Huber, D., Tian, X., Curran, T., O'Reilly, R.C., Woroch, B. (2008). The dynamics of integration and separation: ERP, MEG, and neural network studies of immediate repetition effects. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1389-1416.
- Jacobs, A.M. (2003). Simulative Methoden. In G. Rickheit, T. Herrmann & W. Deutsch (Eds.), Handbuch der Psycholinguistik, (pp. 125-142). Berlin: de Gruyter.
- Jacobs, A.M. (2006). Messung der Hirnaktivität. In J. Funke & P. Frensch (Eds.), Handbuch der Allgemeinen Psychologie, Kognition. Göttingen: Hogrefe.
- Jacobs, A.M. (2008). Kognitive Modellierung und Simulation / „diagram making“. In S. Guggel & M. Herrmann (Hrsg.), Handbuch der Neuro- und Biopsychologie (54 - 60). Göttingen: Hogrefe.
- Jacobs, A.M. (2011). Neurokognitive Poetik: Elemente eines neurokognitiven Modells des literarischen Lesens. (Neurocognitive poetics: Elements of a neurocognitive model of literary reading). In: R. Schrott & A. Jacobs, Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren, München: Hanser.
- Jacobs, A. M., & Carr, T. H. (1995). Mind mappers and cognitive modelers: Toward cross-fertilization. *Behavioral and Brain Sciences*, 18, 362-363.
- Jacobs, A. M., Graf, R., & Kinder, A. (2003). Receiver operating characteristics in the lexical decision task: Evidence for a simple signal-detection process simulated by the multiple read-out model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 29, 481-488.
- Jacobs, A.M., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1311-1334.
- Jacobs, A.M., Hutzler, F. & Engl, V. (2006). Dem Geist auf der Spur: Neurokognitive Methoden zur Messung von Lern- und Gedächtnisprozessen. *Zeitschrift für Erziehungswissenschaft*, 9, 71-86.
- Jacobs, A. M., & Rösler, F. (1999). Dondersian dreams in brain-mappers minds, or, still no cross-fertilization between mind mappers and cognitive modelers. *Behavioral and Brain Sciences*, 22, 293-295.
- Just, M. A., Carpenter, P. A., & Varma, S. (1999). Computational modeling of high-level cognition and brain function. *Human Brain Mapping*. 8, 128-136.
- Klonek, F., Tamm, S., Hofmann, M. J., & Jacobs, A. M. (2009). Does familiarity or conflict account for performance in the word-stem completion task? Evidence from behavioural and event-related-potential data. *Psychological Research*, 73, 871-882.
- Lenartowicz, A., Kalar, D.J., Congdon, E., & Poldrack, R.A. (2010). Towards an

- Ontology of Cognitive Control. *Topics in Cognitive Science*, 2, 678–692
- Lorig, T. S. (2009). What was the question? fMRI and inference in psychophysiology. *International Journal of Psychophysiology*, 73, 17-21.
- Luo, L., Rodriguez, E., Jerbi, K., Lachaux, J.-P., Martinerie, J., Corbetta, M., Shulman, G.L., Piomelli, D., Turrigiano, G. G., Nelson, S. B., Joëls, M., de Kloet, R., Holsboer, F., Amodio, D. M., Frith, C. D., Block, M. L., Zecca, L., Hong, J.-S., Dantzer, R., Kelley, K. W., Craig, A.D. (2010). Ten years of Nature Reviews Neuroscience: insights from the highly cited. *Nature Reviews Neuroscience*, 11, 718-726.
- Massaro, D. W., & Cowan, N. (1993). Information processing models: Microscopes of the mind. *Annual Review of Psychology*, 44, 383-425.
- McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88, 375-407.
- O'Reilly, R. C. & Norman, K. A. (2002). Hippocampal and Neocortical Contributions to Memory: Advances in the Complementary Learning Systems Framework. *Trends in Cognitive Sciences*, 6, 505-510.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+model of reading aloud. *Psychological Review*, 114, 273–315.
- Pitt, M. A., Myung, I.J., Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, 109, 472-491.
- Poeppel, D. (1996). A critical review of PET studies of language. *Brain and Language*, 55, 317 – 351.
- Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Science*, 10, 59–63.
- Poldrack, R.A., Halchenko, Y., & Hanson, S.J., (2009). Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychological Science*, 20, 1364-1372.
- Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology*, 22, 262–275.
- Rey, A., Dufau, S., Massol, S., & Grainger, J. (2008). Testing computational models of letter perception with item-level ERPs. *Cognitive Neuropsychology*, 26, 7-22.
- Schrott, R., & Jacobs, A.M. (2011). Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren, München: Hanser.
- Simon. H. A. (1969). The sciences of the artificial. Cambridge, MA: MIT press.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders'

method. *Acta Psychologica*, 30, 276–315.

