Prüfplan

Version 1, 18.10.2024

Deutscher Titel: Neuronale Signaturen semantischer Gradienten bei der Bildung von falschen Erinnerungen: Eine DRM-spezifische fMRT-Untersuchung

Titel für Probanden: Semantische Netzwerke im Gehirn: Wie wir Worte verarbeiten

Studienleiter

Dr. Gordon Feld
Zentralinstitut für Seelische Gesundheit
Abteilung für klinische Psychologie
J5
68159 Mannheim
Tel.: 0621 / 1703-6540

Synopsis

Semantische Informationen sind im temporalen Pol, dem semantischen Zentrum des Gehirns, gespeichert und lassen sich mittels neurobildgebender Verfahren als differenzierbare neuronale Muster erfassen. Semantische Ähnlichkeiten, basierend auf dem Deese-Roediger-McDermott (DRM) Paradigma, manifestieren sich in den Überlappungen dieser neuronalen Muster und korrelieren mit der Entstehung falscher Erinnerungen. Chadwick et al. (2016) belegten diese Beziehung durch funktionelle Kernspintomographie, indem sie neuronale Aktivierungen während einer Kategorienbeurteilungsaufgabe mit Verhaltensdaten aus DRM-basierten Tests verglichen. Dabei verwendeten sie klassische DRM-Wortlisten (z. B. Apfel, Gemüse, Orange, Kiwi, Zitrusfrucht), wobei jede Liste mit einem spezifischen kritischen ‚Lure', beispielsweise „Obst", assoziiert ist. In dieser Studie untersuchen wir den Gradientencharakter semantischer Repräsentationen sowie den Einfluss verschiedener kategorischer Dimensionen auf die neuronalen Muster semantischer Konzepte und deren Zusammenhang mit der Bildung falscher Erinnerungen. Darüber hinaus analysieren wir, ob die Repräsentationen der semantischen Modelle signifikant mit den neuronalen Aktivierungsmustern korrelieren.

Theoretischer Hintergrund

Falsche Erinnerungen, das Erinnern von Ereignissen, die nie stattgefunden haben, spiegeln die konstruktive Natur des menschlichen Gedächtnisses wider und zeigen den Einfluss semantischer Informationen auf Gedächtnisprozesse. Das Deese-Roediger-McDermott (DRM) Paradigma (Roediger & McDermott, 1995) demonstriert, wie Individuen oft falsche Erinnerungen an oder die Anerkennung von neuen, semantisch verwandten Wörtern haben, die während der Enkodierung präsentiert wurden. Dieses Phänomen hat zu zwei prominenten theoretischen Rahmenwerken geführt: das Activation-Monitoring Framework (AMF) (Roediger et al., 2001) und die Fuzzy Trace Theory (FTT) (Reyna & Brainerd, 1998). AMF schlägt vor, dass falsche Erinnerungen aus der verteilten Aktivierung semantischer Netzwerke während der Enkodierung oder des Abrufs entstehen, gekoppelt mit Fehlern beim Quellenmonitoring. FTT hingegen geht davon aus, dass Erinnerungen sowohl als wörtliche als auch als gist-basierte Spuren enkodiert werden, wobei falsche Erinnerungen entstehen, wenn gist-basierte Repräsentationen ohne spezifische wörtliche Details abgerufen werden. Trotz ihrer konzeptionellen Unterschiede betonen beide Theorien die Rolle semantischer Informationen bei der Bildung falscher Erinnerungen (Gallo, 2010).

Mehrere neurobildgebende Studien haben Belege für diesen semantischen Einfluss geliefert. Chadwick et al. (2016) identifizierten den temporalen Pol als eine Schlüsselregion in diesem Prozess, der als semantischer Knotenpunkt fungiert, wo sich überlappende neuronale Muster aus gelernten Elementen und semantisch verwandten ‚Lures' falsche Erinnerungen hervorrufen können. Der Grad der neuronalen Überlappung im temporalen Pol stimmt mit der Wahrscheinlichkeit falscher Erinnerungen überein. Diese Entdeckungen unterstützen die Idee eines semantischen Ähnlichkeitsgradienten, der Gedächtnisverzerrungen beeinflusst, was wir mit diesem Experiment nachweisen möchten.

Zusätzlich unterstützen rechnerische Modelle und distributionale semantische Modelle (DSMs) wie Wort-Einbettungen dieses Konzept weiter. Gatti et al. (2022) zeigten, dass semantische Ähnlichkeit, abgeleitet aus DSMs, signifikant mit der Auftretenswahrscheinlichkeit falscher Erinnerungen korreliert, wobei eine höhere Ähnlichkeit zwischen kritischen ‚Lures' und DRM-Listenwörtern die Rate falscher Erinnerungen erhöht. Diese Ergebnisse überbrücken verhaltensbezogene Beobachtungen, kognitive Theorien und neuronale Belege und bieten Einblicke in die Beziehung zwischen semantischer Verarbeitung und der Bildung falscher Erinnerungen.

Neuronale Überlappung und falsche Erinnerungen

Die Rolle des temporalen Pols als amodales Zentrum der semantischen Gedächtnisrepräsentation ist gut etabliert. Kurkela und Dennis (2016) fassten in einer Metaanalyse zahlreiche fMRT-Studien zusammen und identifizierten konsistente Aktivierungsmuster, die mit falschen Erinnerungen verbunden sind. Chadwick et al. (2016) erweiterten diese Erkenntnisse, indem sie das Konzept der neuronalen Überlappung einführten und zeigten, dass individuelle Unterschiede in den Repräsentationen im temporalen Pol idiosynkratische Muster falscher Erinnerungen vorhersagen können. Diese Befunde stimmen mit rechnerischen Modellen der semantischen Kognition überein, die den anterioren Temporallappen (ATL), insbesondere den temporalen Pol, als amodales semantisches Zentrum postulieren (Patterson et al., 2007).

Distributional Semantic Models (DSMs)

Distributional Semantic Models (DSMs) haben sich als leistungsstarke Werkzeuge zur Darstellung und Analyse semantischer Beziehungen etabliert, basierend auf der distributionalen Hypothese, dass Wörter in ähnlichen Kontexten ähnliche Bedeutungen haben (Harris, 1954; Firth, 1957). Moderne DSMs wie Word2Vec (Mikolov et al., 2013), fastText (Bojanowski et al., 2017) und GloVe (Pennington et al., 2014) repräsentieren Wörter als dichte Vektoren in hochdimensionalen Räumen, die komplexe semantische und syntaktische Beziehungen erfassen. Diese Modelle integrieren neuronale Netzwerkarchitekturen und fehlergesteuerte Lernmechanismen, die mit psychologisch plausiblen Lernprozessen übereinstimmen (Günther et al., 2019). In der Forschung zu falschen Erinnerungen bieten DSMs einen differenzierten Ansatz zur Quantifizierung semantischer Beziehungen und könnten theoretische Rahmenwerke wie das Activation-Monitoring Framework und die Fuzzy Trace Theory überbrücken, indem sie kontinuierliche Maße sowohl der assoziativen Stärke als auch der gist-basierten Ähnlichkeit bereitstellen (Gatti et al., 2022; Osth et al., 2020).

Referenzen

Aschenbrenner, S., Tucha, O., & Lange, K. W. (2000). Regensburger Wortflüssigkeits-Test: RWT. Hogrefe, Verlag für Psychologie, Göttingen.

Chadwick, M. J., Anjum, R. S., Kumaran, D., Schacter, D. L., Spiers, H. J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proceedings of the National Academy of Sciences*, 113(36), 10180–10185. https://doi.org/10.1073/pnas.1610686113

Gatti, D., Rinaldi, L., Marelli, M., Mazzoni, G., & Vecchi, T. (2022). Decomposing the semantic processes underpinning veridical and false memories. *Journal of Experimental Psychology: General*, *151*(2), 363.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-Space Models of Semantic Representation From a Cognitive Perspective: A Discussion of Common Misconceptions. *Perspectives on Psychological Science*, 14(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Harris, Z. S. (1954). Distributional Structure. *WORD*, 10(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Karamolegkou, A., Abdou, M., & Søgaard, A. (2023). Mapping Brains with Language Models: A Survey. *arXiv preprint arXiv:230X.XXXXX*. https://doi.org/10.18653/v1/2023.findings-acl.618

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language*, 45(1), 160–176. https://doi.org/10.1006/jmla.2000.2771

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*. University of Illinois Press.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976–987. https://doi.org/10.1038/nrn2277

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162

Reyna, V. F., & Brainerd, C. J. (1998). Fuzzy-Trace Theory and False Memory: New Frontiers. *Journal of Experimental Child Psychology*, 71(2), 194–209. https://doi.org/10.1006/jecp.1998.2472

Roediger III, H. L., Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. In *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). American Psychological Association. https://doi.org/10.1037/10394-006

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803–814.

Schacter, D. L. (2001). Memory distortion: History and current status. In D. L. Schacter,J. T. Coyle, G. D. Fischbach, M. M. Mesulam, & L. E. Sullivan (Eds.), *Memory Distortion*. Cambridge, MA: Harvard University Press.

Hypothesen

1. **Neuronale Überlappung und falsche Erinnerungen:** Größere neuronale Überlappung zwischen DRM-Listenwörtern und ihren assoziierten ‚Lures' im temporalen Pol korreliert positiv mit den Raten falscher Erinnerungen sowohl in Erkennungs- als auch in Abrufaufgaben.

2. **Gedächtnisdifferenzierung und Antwortverzerrung:**
   Größere neuronale Überlappung zwischen DRM-Listenwörtern und ‚Lures' ist mit einer verringerten Diskriminierungsfähigkeit und einer erhöhten Verzerrung hin zu der Annahme von Elementen als „alt" in Erkennungsaufgaben verbunden.
3. **Gedächtnisentscheidungsprozesse:**
   Erhöhte neuronale Überlappung beeinflusst die Geschwindigkeit und Vorsicht bei Gedächtnisentscheidungen, wobei eine schnellere Akkumulation von Beweisen zugunsten von „alt"-Antworten und weniger vorsichtiges Entscheidungsverhalten erwartet wird.
4. **Neuronale Ausrichtung mit distributionalen semantischen Modellen (DSMs):**
   Neuronale Repräsentationen von Wörtern im temporalen Pol zeigen signifikante Korrelationen mit Repräsentationsähnlichkeitsmatrizen, die aus distributionalen semantischen Modellen abgeleitet wurden, wobei Repräsentation im besser ausgerichtet sind.
5. **Semantische Differentialbewertungen und neuronale Muster:**
   Wörter, die in semantischen Differentialdimensionen ähnlicher bewertet werden, zeigen eine größere neuronale Musterähnlichkeit im temporalen Pol, und kritische ‚Lures', die semantisch ähnlicher zu ihren assoziierten DRM-Listenwörtern sind, produzieren häufiger falsche Erinnerungen.

Versuchsablauf

Eine Woche vor Beginn des Experiments werden die Teilnehmer zunächst schriftlich und mündlich über den Versuch informiert. Sie unterschreiben dann die Einverständniserklärung, nachdem sie die Möglichkeit hatten, Fragen zu stellen. Anschließend geben sie ihre demographischen Daten an.

Diese Studie umfasst zwei Sitzungen im Abstand von mindestens 21 Tagen. Die erste Online-Sitzung beinhaltet verschiedene Aufgaben: eine Psychomotorische Vigilanzaufgabe zur Bewertung der Aufmerksamkeit und Reaktionszeiten, die Deese-Roediger-McDermott (DRM) falsche Erinnerung Aufgabe mit Enkodierungsphasen, einem Hagen-Matrizen-Test zur Ablenkung, freiem Abruf und Wiedererkennungstests, die Regensburger Wortflüssifgkeitstest zur Bewertung der verbalen Flüssigkeit sowie eine Semantische Differential Aufgabe zur Messung semantischer Urteile von Wörtern. Die zweite Sitzung umfasst einen funktionalen Magnetresonanztomographie (fMRT) Scan, bei dem den Teilnehmern die Ziel- und ‚Lure'-Wörter aus der DRM-Wiedererkennungstest präsentiert werden, während sie eine beiläufige Wortklassifikationsaufgabe durchführen.

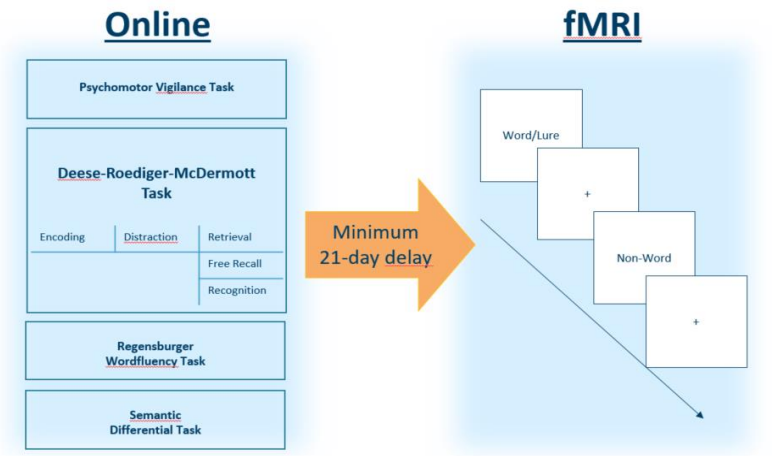

Abbildung 1. *Experimenteller Ablauf.*

Psychomotor-Vigilanz-Aufgabe(PVT)

Um die Aufmerksamkeit der Teilnehmer während des gesamten Experiments zu gewährleisten, werden wir eine 3-minütige Version der Psychomotorischen Vigilanzaufgabe (Roach et al., 2006) durchführen. Die Teilnehmer werden angewiesen, die Leertaste so schnell wie möglich zu drücken, wenn eine Millisekundenuhr auf dem Bildschirm erscheint.

Hagen-Matrizen-Test

Der Hagen-Matrizen-Test (lange Form) ist eine Aufgabe zur Messung der fluiden Intelligenz und abstrakten Denkfähigkeiten (Heydash & Tsantidis, 2014). Er besteht aus einer Reihe zunehmend komplexer visueller Matrizen, bei denen den Teilnehmern ein 3x3-Raster abstrakter Muster mit einem fehlenden Stück präsentiert wird. Die Aufgabe besteht darin, das korrekte fehlende Stück aus acht Optionen auszuwählen, um das Muster logisch zu vervollständigen. Die Teilnehmer müssen innerhalb eines Zeitlimits von 40 Minuten 20 Matrizen lösen. In dieser Studie dient der Hagen-Matrizen-Test primär als non-verbale Ablenkungsaufgabe zwischen den Enkodierungs- und Abrufphasen des DRM-Paradigmas.

Deese Roediger McDermott Aufgabe (DRM)

1. Enkodierungsphase:

    - Teilnehmern werden DRM-Wortlisten präsentiert, jede bestehend aus semantisch verwandten Wörtern. Basierend auf Chadwick et al. (2016) und den Erkenntnissen von McDermott und Watson (2001) werden die Wörter visuell auf einem Computerbildschirm jeweils für 500 ms angezeigt, mit einem 3-Sekunden-Intervall zwischen den Listen. Innerhalb jeder Liste werden die Wörter in abnehmender Reihenfolge der assoziativen Stärke zum kritischen ‚Lure' präsentiert. Die Reihenfolge der Listenpräsentation

wird für jeden Teilnehmer randomisiert. Teilnehmer werden angewiesen, die Wörter für einen späteren Gedächtnistest zu memorieren.

2. Hagen Matrizen:

    - Nach der Enkodierungsphase führen die Teilnehmer eine Ablenkungsaufgabe durch, um aktives Wiedererinnern der gelernten Wörter zu verhindern. Diese Aufgabe ist der Hagen-Matrizen-Test (lange Form).

3. Freier Abruf:

    - Unmittelbar nach der Ablenkungsphase haben die Teilnehmer 10 Minuten Zeit, sich frei an so viele Wörter wie möglich aus der Enkodierungsphase zu erinnern. Sie geben ihre Antworten in ein Textfeld auf dem Computerbildschirm ein. Nach jedem eingegebenen Wort bewerten die Teilnehmer ihr Vertrauen in das Vorhandensein des Wortes während der Enkodierung auf einer 4-Punkte-Skala (1 = "sehr unsicher" bis 4 = "sehr sicher").

4. Wiedererkennungsphase:

    - Nach dem freien Abruf absolvieren die Teilnehmer einen schnellen Wiedererkennungstest, der 380 Wörter umfasst (160 Zielwörter, 160 nicht verwandte Wörter, 60 ‚Lures'). Die Wörter werden einzeln in zufälliger Reihenfolge präsentiert. Für jedes Wort müssen die Teilnehmer innerhalb eines 5-Sekunden-Fensters eine schnelle Entscheidung zwischen "alt" und "neu" treffen.

      Erfolgt innerhalb dieses Zeitfensters keine Reaktion, wird der Versuch als Fehlversuch gewertet und das nächste Wort präsentiert. Nach jeder "alt"/"neu"-Entscheidung geben die Teilnehmer eine Vertrauensbewertung auf einer 4-Punkte-Skala ab, ohne zeitliche Begrenzung für diese Bewertung.

Semantische Differential Aufgabe

Teilnehmer absolvieren eine semantische Differentialaufgabe für jedes präsentierte Wort, einschließlich der Ziel- und ‚Lure'-Wörter. Sie bewerten jedes Wort auf sechs Dimensionen, die aus dem semantischen Differential von Osgood et al. (1957) abgeleitet sind: gut-schlecht, stark-schwach, aktiv-passiv, abstrakt-konkrett, vertraut-unvertraut und leicht zu merken-schwer zu merken. Diese Dimensionen werden in zufälliger Reihenfolge präsentiert, um Antwort-Habituation zu verhindern.

Teilnehmer bewerten jedes Wort auf einer kontinuierlichen Skala von 0 bis 100 mithilfe eines Schiebereglers, was nuancierte und individualisierte Antworten ermöglicht.

Regensburger Wortfluency Task (RWT)

Am Ende der Verhaltenssitzung führen die Teilnehmer eine Aufgabe zur Bewertung ihrer verbalen Flüssigkeit durch (Aschenbrenner et al., 2000). Die Teilnehmer werden gebeten, innerhalb von 2 Minuten so viele Wörter wie möglich zu bilden, die mit dem Buchstaben "m" oder "p" beginnen. Die Reihenfolge der Buchstaben wird zwischen den Teilnehmern randomisiert, um Reihenfolge-Effekte zu kontrollieren.

fMRT Messung

Mindestens drei Wochen nach der Verhaltenssitzung werden die Teilnehmer einem fMRT-Scan unterzogen. Die fMRT-Aufgabe besteht aus sechs funktionalen Durchläufen, bei denen die 160 Wörter der DRM-Liste und 60 ‚Lures' einzeln für jeweils 3 Sekunden präsentiert werden. Jedes Wort wird mehrfach präsentiert, um stabile neuronale Musterschätzungen zu gewährleisten. Die Reihenfolge der Wortpräsentation wird randomisiert, wobei die Wörter aus derselben DRM-Liste nie nacheinander erscheinen. Während der Präsentation führen die Teilnehmer eine beiläufige Aufgabe durch, bei der sie eine Taste drücken, wenn ein Nicht-Wort erscheint.

Die Bildgebung erfolgt mit einem 7 Tesla (7T) MRI-Scanner, der eine hohe räumliche Auflösung und verbessertes Signal-Rausch-Verhältnis bietet. Die Daten werden mittels standardisierter neuroimaging-Pipelines vorverarbeitet und für die Representational Similarity Analysis (RSA) vorbereitet.

Proband*innen

Wir werden insgesamt n = 44 Teilnehmer (im Alter von 18-35 Jahren) rekrutieren. Um die angemessene Stichprobengröße für unsere Studie zu bestimmen, führten wir eine a priori Power-Analyse mit der Software G*Power 3.1 (Faul et al., 2009) durch. Unsere primäre Analyse beinhaltet die Bewertung der Korrelation zwischen neuronaler Überlappung in dem Temporalpol und der Rate falscher Erinnerungen, wie in Hypothese H1 dargelegt. Basierend auf Chadwick et al. (2016), die eine Korrelation von r = 0,40 zwischen neuronaler Überlappung und der Wahrscheinlichkeit falscher Erinnerungen über DRM-Wortlisten berichteten, verwendeten wir diese Effektgröße als Schätzung für die erwartete Beziehung in unserer Studie.

Während Chadwick et al. (2016) ihre Studie mit 18 Teilnehmern durchführten, beabsichtigen wir, eine ausreichende statistische Power sicherzustellen, um die hypothesierten Effekte zuverlässig nachweisen zu können. Unter Verwendung der Parameter eines zweiseitigen Tests, eines Alpha-Niveaus (α) von 0,05, einer gewünschten Power (1 - β) von 0,80 und einer Effektgröße von r = 0,40 ergab die Power-Analyse eine erforderliche Stichprobengröße von 44 Teilnehmern. Diese Berechnung berücksichtigt potenzielle Variabilität in den Effektgrößen und entspricht den aktuellen Standards, die ausreichend leistungsstarke neuroimaging Studien betonen, um die Reproduzierbarkeit und Zuverlässigkeit zu erhöhen. Diese Stichprobengröße ermöglicht es uns, eine signifikante Kor-

relation zwischen neuronaler Überlappung und der Rate falscher Erinnerungen mit einer Power von 80 % nachzuweisen, wodurch sichergestellt wird, dass unsere Ergebnisse sowohl statistisch robust als auch mit früheren Forschungen in diesem Bereich vergleichbar sind.

Die Teilnehmenden werden eingeschlossen, wenn sie zwischen 18 und 35 Jahre alt sind.

Sie sollen deutsche Muttersprachler, sowie Rechtshänder sein und über ein normales (oder zu normal korrigiertes) Sehvermögen verfügen.

Ausgeschlossen werden Teilnehmende mit (I) Erkrankungen und/oder Einnahme von Medikamenten und/oder Drogen, die das Nervensystem und/oder die Lernfähigkeit beeinträchtigen, (II) Schwangere, sowie (III) Probanden, welche die Kriterien der PVT Reaktionszeit (<100 ms oder >1000 ms in ≥ 20% der Aufgaben) nicht erfüllen.

MRT-Ausschlusskriterien: Metall im Körper (z.B. Herzschrittmacher oder Schrauben), Metallverarbeitende Berufe oder Hobbys, großflächige Tattoos, nicht-entfernbarer Metallschmuck, Angst vor engen Räumen.

Datenhaltung, Ort, Verantwortliche Stelle, Dauer der Speicherung

ZI Mannheim, Abteilung Klinische Psychologie, Dr. Gordon Feld, 10 Jahre

Kodierungsart (Doppelte Kodierung, Zugangsrechte, Dekodierung im Notfall möglich?)

Doppelkodierung, Zugang nur durch Projektmitarbeiter, Dekodierung im Notfall während der Projektlaufzeit möglich.

Datenweitergabe von pseudonymisierten Daten an Dritte

Wenn die Ergebnisse publiziert werden, werden die pseudonymisierten Forschungsdaten entsprechend den Vorgaben der guten wissenschaftlichen Praxis 10 Jahre lang gespeichert und auf Verlangen zur Überprüfung der Ergebnisse anderen Forschern zugänglich gemacht. Die Daten werden auf einem zentralen Server gespeichert, der durch die IT-Abteilung des ZI Mannheim betreut wird. Darüberhinaus werden die anonymisierten Daten auf einem Onlinerepositorium mit Server in Deutschland (PsychArchives des Leibniz-Institut für Psychologie (ZPID)) frei zugänglich gemacht. Für MRT Daten werden nur die First-Level-Beta-Gewichte und die Con-Images online geteilt, da es ansonsten trotz Anonymisierung ein zu hohes Re-Identifizierungsrisikogäbe.